NON-NORMAL BIVARIATE DISTRIBUTIONS: ESTIMATION AND HYPOTHESIS TESTING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SAHAR BOTROS QUMSIYEH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
STATISTICS

OCTOBER 2007

Approval of the thesis:

NON-NORMAL BIVARIATE DISTRIBUTIONS: ESTIMATION AND HYPOTHESIS TESTING

submitted by **SAHAR QUMSIYEH** in partial fulfillment of the requirements of the degree of **Doctor of Philosophy in Statistics Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen ________________
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Ali Uzun ________________
Head of Department, **Statistics**

Prof. Dr. Moti Tiku ________________
Supervisor, **Statistics Dept., METU**

Assoc. Prof. Dr. Qamarul Islam ________________
Co-Supervisor, **Economics Dept., ÇÜ**

Examining Committee Members:

Assoc. Prof. Dr. Birdal Şenoğlu ________________
Statistics Dept., AÜ

Prof. Dr. Moti Tiku ________________
Statistics Dept., METU

Assoc. Prof. Dr. Ayşen Akkaya ________________
Statistics Dept., METU

Dr. Barış Sürücü ________________
Statistics Dept., METU

Prof. Dr. G. Wilhelm Weber ________________
Institute of Applied Mathematics, METU

Date: ________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Sahar Qumsiyeh

Signature :

ABSTRACT

NON-NORMAL BIVARIATE DISTRIBUTIONS:
ESTIMATION AND HYPOTHESIS TESTING

Qumsiyeh, Sahar

Ph. D., Department of Statistics

Supervisor : Prof. Dr. Moti Lal TIKU

Co-Supervisor: Assoc. Prof. Dr. Qamarul ISLAM

October 2007, 262 pages

When using data for estimating the parameters in a bivariate distribution, the tradition is to assume that data comes from a bivariate normal distribution. If the distribution is not bivariate normal, which often is the case, the maximum likelihood (ML) estimators are intractable and the least square (LS) estimators are inefficient. Here, we consider two independent sets of bivariate data which come from non-normal populations. We consider two distinctive distributions: the marginal and the conditional distributions are both Generalized Logistic, and the marginal and conditional distributions both belong to the Student's t family. We use the method of modified maximum likelihood (MML) to find estimators of various parameters in each distribution. We perform a simulation study to show that our estimators are more efficient and robust than the LS estimators even for small sample sizes.

We develop hypothesis testing procedures using the LS and the MML estimators. We show that the latter are more powerful and robust. Moreover, we give a comparison of our tests with another well known robust test due to Tiku and Singh (1982) and show that our test is more powerful. The latter is based on censored normal samples and is quite prominent (Lehmann, 1986). We also use our MML estimators to find a more efficient estimator of Mahalanobis distance. We give real life examples.

Key Words: Modified maximum likelihood (MML), Least squares (LS), Bivariate data, Non-normal error distribution, Robustness.

ÖZ

NORMAL OLMAYAN İKİDEĞİŞKENLİ DAĞILIMLAR: TAHMİN VE HİPOTEZ TESTİ

Qumsiyeh, Sahar
Doktora, İstatistik Bölümü
Tez Yöneticisi : Prof. Dr. Moti Lal TİKU
Ortak Tez Yöneticisi: Yrd. Doç. Dr. Qamarul İSLAM

Ekim 2007, 262 sayfa

İkideğişkenli bir dağılımın parametlerinin tahmini için kullanılan geleneksel yöntem, dağılımın ikideğişkenli normal dağılımdan geldiği varsayımı altında bu tahminleri yapmaktır. Ancak, gerçekte veriler iki değişkenli normal dağılımdan gelmiyor ise, En Çok Olabilirlik tahmin edicilerinin elde edilmesi zordur. Bu durumda bulunan En Küçük Kareler hata tahmin edicileri ise etkin tahmin ediciler değildirler. Biz, bu çalışmada normal kitleden gelmeyen ikideğişkenli iki bağımsız veri kümesini ele alıp, elimizde iki tane rasgele vektör olduğunu varsayalım. Öncelikle, hem ilk vektör hem de ikinci vektör için marjinal olasılık yoğunluk fonksiyonu ve koşullu olasılık yoğunluk fonksiyonları olarak Genelleştirilmiş Lojistik dağılımını ele aldık. Daha sonra ise, hem ilk vektör hem de ikinci vektör için marjinal olasılık yoğunluk fonksiyonu ve koşullu olasılık yoğunluk fonksiyonları olarak Student t dağılımını ele aldık. Ele alacağımız modellerdeki parametrelerin tahmin edicilerini bulmak için Uyarlanmış En Çok Olabilirlik tahmin yöntemini kullandık. Elde ettiğimiz tahmin edicilerin, örneklem büyüklüğü küçük iken bile En küçük kareler tahmin edicilerinden daha etkin ve güçlü olduklarını gösterebilmek için bir benzetim çalışması yaptık.

En küçük kareler ve Uyarlanmış En Çok Olabilirlik tahmin edicileri için hipotez testleri geliştirip, Uyarlanmış En Çok Olabilirlik yöntemiyle elde edilen test istatistiklerinin daha güçlü ve sağlam olduklarını gösterdik. Ayrıca, elde edeceğimiz bu yeni testlerin, Tiku ve Singh tarafından 1982 yılında önerilen ve güçlü olduğu bilinen test istatistiği ile karşılastırdık ve bu çalışmada önerdiğimiz testlerin Tiku ve Singh'in testinden daha güçlü olduğunu gösterdik. Aynı zamanda, bulduğumuz Uyarlanmış En Çok Olabilirlik tahmin edicilerini, Mahalanobis uzaklığının daha etkin bir tahmin edicisini bulmak için de kullandık. Ayrıca, gerçek hayat verileri üzerinde uygulama yaptık.

Anahtar Kelimeler: Uyarlanmış En Çok Olabilirlik (UEÇO), En küçük kareler (EKK), İkideğişkenli veri, Normal olmayan hata dağılımı, Sağlamlık.

To My Parents

ACKNOWLEDGMENTS

The author wishes to express her appreciation to her supervisor Prof. Dr. Moti L. Tiku for all his support, help, encouragement, insight and guidance.

The author would like to thank Assoc. Prof. Dr. Qamarul Islam for his suggestions and comments.

The author also thanks the Turkish government for providing the financial support needed. It was highly appreciated.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND COMPARISON OF METHODS

Consider a location-scale family of distributions, $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, where μ is the location parameter and σ is the scale parameter; μ and σ may or may not be the mean and standard deviation of the population. Obtaining efficient, unbiased and robust estimators of μ and σ is of paramount importance. Many methods of estimation have been introduced in the literature. Three prominent methods are: Maximum likelihood (ML), Modified Maximum Likelihood (MML), and Least Squares (LS). In this chapter, we summarize the above three methods and discuss the merits and demerits of each. We give a brief explanation of how each method finds estimators of μ and σ. After this general explanation of each method we consider, as examples, three distributions and show how each method is used to find estimators of μ and σ. The families of distributions we consider are: Student's t, Generalized Logistic and Short Tail symmetric. The procedure developed for each of these families can be extended to estimate a shape parameter also (Tiku and Akkaya, 2004).

1.1 Maximum Likelihood (ML)

This has been the most widely used method of estimation. It is based on finding estimators which maximize the likelihood function. It relies on the fact that the full functional form of the distribution is known. It proceeds as follows:

Suppose we have one of a two parameter family as the distribution of X. The likelihood function is

$$L = \prod_{i=1}^{n} f(x_i, \mu, \sigma).$$

Taking natural logarithm we get

$$\ln L = \sum_{i=1}^{n} \ln(f(x_i, \mu, \sigma)).$$

Since lnL has one-to-one correspondence with L, maximizing L is equivalent to maximizing lnL. To find the estimators $\hat{\mu}$ and $\hat{\sigma}$ which maximize lnL, we differentiate lnL with respect to μ and σ. We obtain the following equations which are called the Maximum Likelihood equations:

$$\frac{\partial \ln L}{\partial \mu} = 0 \text{ and} \tag{1.1.1}$$

$$\frac{\partial \ln L}{\partial \sigma} = 0. \tag{1.1.2}$$

The solutions of these equations are the maximum likelihood estimators (MLE). If the equations do not have explicit solutions, we must solve them by iteration to find the MLE of μ and σ.

The MLE of μ and σ have many desirable properties. They are consistent for μ and σ and BAN (best asymptotically normal) under some very general regularity conditions. The MLE are functions of jointly sufficient statistics for μ and σ if the latter exists. Under some regularity conditions, the MLE achieve minimum variance (as given by the Cramer-Rao lower bound) at any rate asymptotically. Another desirable property of the MLE is the invariance property. This means, for example, if $\hat{\mu}$ is the MLE of μ then the MLE of $\alpha = h(\mu)$ is $\hat{\alpha} = h(\hat{\mu})$, provided $h(\mu)$ is a monotonic function of μ.

Because of the above properties, the maximum likelihood method has been the most widely used method of estimation. However, this method has drawbacks. The MLE can have substantial bias at any rate for small n (sample size). Therefore, when using the ML method of estimation we must make sure we correct for bias. In addition to that, often, the maximum likelihood equations don't have explicit solutions and we must use iterative methods to solve them. Solving the equations by iteration is problematic because iterations may not converge and, if they do, they may converge to wrong values. The presence of outliers and inliers in the data can adversely affect the convergence of these iterative solutions (Puthenpura & Sinha, 1986). Difficulties especially arise when we have many parameters which means there are too many equations to iterate simultaneously. The iterative solutions are often unreliable.

1.2 Modified Maximum Likelihood (MML)

The reason that the maximum likelihood equations cannot be solved in most situations is the presence of non-linear functions which makes the equations intractable. The Modified Maximum Likelihood method approximates these functions by linear functions in such a way that the differences between the two converge to zero as n tends to infinity. To find estimators of μ and σ, the MML method proceeds as follows:

1. We order the x_i's in ascending order of magnitude so that

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$$

2. We express equations (1.1.1) and (1.1.2) in terms of the order statistics $x_{(i)}$ ($1 \le i \le n$). This does not result in any change in the numerical values of the equations since complete sums are invariant to ordering. We also express the equations in terms of the standardized variates $z_{(i)} = \frac{x_{(i)} - \mu}{\sigma}$. Note that $z_{(i)}$'s have the same order as the $x_{(i)}$'s since μ is a constant and σ is positive.

3. The nonlinear functions in the ML equations are approximated by using the first two terms of their Taylor series expansion around $t_{(i)} = E(z_{(i)})$, $1 \le i \le n$; $t_{(i)}$'s can be found (approximated) from the following equation:

$$\int_{-\infty}^{t_{(i)}} f(z)dz = \frac{i}{n+1}$$, where f is the p.d.f. of Z.

4. Replacing the nonlinear functions in equations (1.1.1) and (1.1.2) by linear approximations, we get the following modified maximum likelihood equations

$$\frac{\partial \ln L^*}{\partial \mu} = 0 \text{ and} \tag{1.2.1}$$

$$\frac{\partial \ln L^*}{\partial \sigma} = 0. \tag{1.2.2}$$

which have an explicit solution. The solutions are called the modified maximum likelihood estimators (MMLE) of μ and σ.

Now, since under some very general regularity conditions $\lim_{n\to\infty}\frac{1}{n}\left\{\frac{\partial \ln L^*}{\partial \mu}-\frac{\partial \ln L}{\partial \mu}\right\}=0$ and $\lim_{n\to\infty}\frac{1}{n}\left\{\frac{\partial \ln L^*}{\partial \sigma}-\frac{\partial \ln L}{\partial \sigma}\right\}=0$ (Tiku and Akkaya, 2004), we see that the modified likelihood equations are asymptotically equivalent to the corresponding maximum likelihood equations; for a rigorous proof, see Vaughan and Tiku (2000). This means that the MMLE have, asymptotically, the same desirable properties as the MLE. Namely, under some very general regularity conditions they are consistent, BAN, and efficient. Unlike equations (1.1.1) and (1.1.2), the MML equations (1.2.1) and (1.2.2) always have an explicit solution and are, therefore, computationally viable. In addition, the MMLE are self bias correcting, whereas the MLE sometimes have substantial bias. Another advantage of this method is that the MMLE are robust to the presence of outliers in the data and other data anomalies.

The drawback of this method is that it cannot be applied to every distribution, e.g. Tukey lambda-family (Akkaya and Tiku, 2005). However, no scientific method is universally applicable.

1.3 Least Squares

The least squares technique is based on minimizing the error sum of squares. If we write

$x_i = \mu + e_i, \; 1 \le i \le n,$

the least squares method finds the estimator of μ by minimizing the error sum of squares $\sum_{i=1}^{n} e_i^2$. The LS estimator of σ^2, $\tilde{\sigma}^2$, is formulated to be

$\tilde{\sigma}^2 = \frac{1}{(n-1)} \min\left[\sum_{i=1}^{n} e_i^2\right]$ (Gauss norm).

The advantage of this method is that we do not need to make any assumption about the distribution in order to find the least square estimators (LSE). If the errors are iid normal with mean zero and constant variance then the LSE are the same as the MLE and are fully efficient. However, if the distribution is not normal, the LSE loose efficiency. The least square estimators also loose efficiency relative to MMLE as n increases. The LSE need to be adjusted for bias if the errors have a non-zero mean, or the variance is not σ^2. Another disadvantage of the LSE is that they are not at all robust to the presence of data anomalies.

1.4 Examples

1.4.1 Generalized Logistic

Here, the p.d.f. has the form

$$f(x) = \frac{b}{\sigma} \frac{\exp\left(-\frac{x-\mu}{\sigma}\right)}{\left[1+\exp\left(-\frac{x-\mu}{\sigma}\right)\right]^{b+1}}. \qquad (1.4.1)$$

Suppose we have a random sample from the above distribution. We are interested in estimating μ and σ.

(1) *Maximum Likelihood*

We write the likelihood function

$$L = \prod_{i=1}^{n} \left[\frac{b}{\sigma} \frac{\exp\left(-\dfrac{x_i - \mu}{\sigma}\right)}{\left[1 + \exp\left(-\dfrac{x_i - \mu}{\sigma}\right)\right]^{b+1}} \right].$$

Letting $z_i = \dfrac{x_i - \mu}{\sigma}$ and taking the natural log of the likelihood function we get

$$\ln L \propto -n \ln \sigma - \sum_{i=1}^{n} z_i - (b+1) \sum_{i=1}^{n} \ln\left(1 + e^{-z_i}\right). \tag{1.4.2}$$

We now differentiate the above equation with respect to μ and σ and get the following equations (called the maximum likelihood equations):

$$\frac{\partial \ln L}{\partial \mu} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma} \sum_{i=1}^{n} \frac{e^{-z_i}}{(1 + e^{-z_i})} = 0 \text{ and} \tag{1.4.3}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^{n} z_i - \frac{(b+1)}{\sigma} \sum_{i=1}^{n} \frac{z_i e^{-z_i}}{(1 + e^{-z_i})} = 0. \tag{1.4.4}$$

Equations (1.4.3) and (1.4.4) don't have explicit solutions. The equations must be solved by iteration.

Example 1.4.1

We generated n = 100 random numbers from the distribution (1.4.1) assuming b = 1, μ = 0 and σ = 1. We took equations (1.4.3) and (1.4.4) and solved them by iterations. The method we used was the modified Powell hybrid algorithm. This algorithm is a variation of Newton's method, which uses a finite-difference approximation to the Jacobian and takes precautions to avoid large step sizes or increasing residuals. For further description, see More et al. (1980). The results of the iterations are given in Table 1.4.1. The first two columns in the table represent the numerical value of equation (1.4.3) and equation (1.4.4) at each iteration step. The last two columns represent the current values of $\hat{\mu}$ and $\hat{\sigma}$ obtained with each iteration. The values -1 and 5 are the initial values to initiate the iterations.

Table 1.4.1 ML equations—iteration results.

Eq 1	Eq 2	$\hat{\mu}$	$\hat{\sigma}$
8.40	-92.25	-1.00	5.00
8.40	-92.26	-1.00	5.00
8.40	-92.26	-1.00	5.00
9.22	-98.15	6.07	-33.45
13.96	-95.78	-6.79	23.67
8.40	-92.26	-1.00	5.00
8.40	-92.26	-1.00	5.00
9.22	-98.15	6.07	-33.45
13.96	-95.78	-6.79	23.67
-5.49	-92.45	-0.67	-4.77
-28.30	-80.78	-4.63	-7.64
-99.99	1028.40	-11.84	-1.04
-23.85	-76.12	-1.96	-3.54
-10.50	36.74	0.17	0.86
-33.08	-45.27	1.56	2.13
-8.53	2211.85	0.12	0.06
-10.50	36.74	0.17	0.86
-10.50	36.68	0.17	0.86
-0.56	7.54	-0.13	1.00
-0.12	2.08	-0.14	1.04
0.00	0.14	-0.14	1.05
0.00	0.01	-0.14	1.05
0.00	0.00	-0.14	1.05

The iterations converged giving the following solutions: $\hat{\mu}$ = -0.145 and $\hat{\sigma}$ = 1.05. This seems to be a good solution; it is close enough to the true values of μ and σ. We now include outliers in the data, changing r of the x's to be outliers (r = [0.1n + 0.5]). Thus, $n - r$ of the *x's* come from the distribution with p.d.f. given in equation (1.4.1) with mean μ and variance σ^2 and r of the *x's* (we don't know which) come from the same distribution with mean μ and variance $(4\sigma)^2$. The results of the iterations are given in Table 1.4.2.

Table 1.4.2 ML equations—iteration results, outlier model.

Eq 1	Eq 2	$\hat{\mu}$	$\hat{\sigma}$
9.63	-82.34	-1.00	5.00
9.63	-82.34	-1.00	5.00
9.63	-82.35	-1.00	5.00
11.06	-93.92	2.77	-11.81
20.87	-88.01	-5.26	12.49
9.63	-82.34	-1.00	5.00
9.63	-82.35	-1.00	5.00
11.06	-93.92	2.77	-11.81
20.87	-88.01	-5.26	12.49
59.10	191.03	-2.21	0.87
-3.82	-68.34	0.26	3.25
-14.06	38.82	0.39	1.17
3.20	-48.64	-0.24	2.32
3.56	-23.74	-0.25	1.76
1.65	-7.74	-0.17	1.53
-0.47	2.77	-0.09	1.42
-0.01	-0.20	-0.11	1.45
0.01	0.00	-0.11	1.45
0.00	0.00	-0.11	1.45

The iterations converge again giving the following solutions:
$\hat{\mu}$ = -0.106 and $\hat{\sigma}$ = 1.445.

Note that the true value of σ here is 1.58, which means both estimators have a bias of about 0.1 even though the sample size is relatively large (n = 100).

Now we include stronger outliers in the data. Suppose *n* - *r* of the *x's* come from the distribution with p.d.f. given in equation (1.4.1) with mean μ and variance σ^2 and *r* of the *x's* (we don't know which) come from the same distribution with mean μ and variance $(12\sigma)^2$. In this case, the iterations don't converge at all. At the fourth iteration the calculated value of the ML equations become infinity as shown in Table 1.4.3.

Table 1.4.3 ML equations—iteration results, strong outliers.

Eq 1	Eq 2	$\hat{\mu}$	$\hat{\sigma}$
10.16	-52.14	-1.00	5.00
10.16	-52.14	-1.00	5.00
10.16	-52.16	-1.00	5.00
-Infinity	Infinity	0.88	-0.08

(2) *Modified Maximum Likelihood*

The reason that equations (1.4.3) and (1.4.4) do not have explicit solutions is the presence of the intractable function

$$g(z) = \frac{e^{-z}}{(1+e^{-z})}.$$

The modified maximum likelihood method proceeds as follows:

First we order the x_i's and express equations (1.4.3) and (1.4.4) in terms of the order statistics as follows

$$\frac{\partial \ln L}{\partial \mu} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} g(z_{(i)}) = 0 \text{ and} \tag{1.4.5}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_{(i)} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} z_{(i)} g(z_{(i)}) = 0, \tag{1.4.6}$$

where $z_{(i)} = \dfrac{x_{(i)} - \mu}{\sigma}$. Note that $z_{(i)}$'s have the same order as the $x_{(i)}$'s since μ is constant and σ is positive.

We approximate the function $g(z_{(i)}) = \dfrac{e^{-z_{(i)}}}{(1+e^{-z_{(i)}})}$ by the first two terms of its Taylor expansion around $t_{(i)} = E(Z_{(i)})$:

$g(z_{(i)}) \cong \alpha_i - \beta_i z_{(i)}$.

To determine the values of alphas and betas we write:

$g(z_{(i)}) \cong g(t_{(i)}) + (z_{(i)} - t_{(i)})g'(t_{(i)})$.

$$g(z_{(i)}) \cong \frac{e^{-t_{(i)}}}{(1+e^{-t_{(i)}})} + (z_{(i)} - t_{(i)})\left(\frac{-e^{-t_{(i)}}}{(1+e^{-t_{(i)}})^2}\right).$$

Thus,

$$\beta_i = \frac{e^{-t_{(i)}}}{(1+e^{-t_{(i)}})^2} \text{ and } \alpha_i = \frac{e^{-t_{(i)}}}{(1+e^{-t_{(i)}})} + \beta_i t_{(i)}.$$

Here $t_{(i)}$'s are found from the following equation

$t_{(i)} = -\ln\left(q_i^{-1/b} - 1\right)$, where $q_i = \dfrac{i}{n+1}$.

We replace the function g by its linear approximation and obtain the following modified maximum likelihood equations:

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n}(\alpha_i - \beta_i z_{(i)}) = 0 \text{ and} \tag{1.4.7}$$

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_{(i)} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} z_{(i)}(\alpha_i - \beta_i z_{(i)}) = 0. \tag{1.4.8}$$

Unlike equations (1.4.3) and (1.4.4), equations (1.4.7) and (1.4.8) have explicit solutions which are the following MMLE:

$$\hat{\mu} = K - D\hat{\sigma} \quad \text{where}$$

$$K = \frac{1}{m}\sum_{i=1}^{n}\beta_i x_{(i)} \text{ and } D = \frac{1}{m}\sum_{i=1}^{n}(\alpha_i - (b+1)^{-1}) \quad (m = \sum_{i=1}^{n}\beta_i).$$

To get an estimator of σ, we solve equation (1.4.8). Rearranging , equation (1.4.8) can be written as

$$-n\sigma^2 - \sigma(b+1)\left[\sum_{i=1}^{n}(\alpha_i - (b+1)^{-1})x_{(i)} - mD\mu\right] + (b+1)\left[\sum_{i=1}^{n}\beta_i x_{(i)}^2 - 2mK\mu + m\mu^2\right] = 0. \tag{1.4.9}$$

Replacing μ by $\hat{\mu} = K - D\hat{\sigma}$, equation (1.4.9) reduces to

$$n\hat{\sigma}^2 + \hat{\sigma}(b+1)\sum_{i=1}^{n}(\alpha_i - (b+1)^{-1})(x_{(i)} - K) - (b+1)\sum_{i=1}^{n}\beta_i(x_{(i)} - K)^2 = 0.$$

Solving the above quadratic equation we get

$$\hat{\sigma} = \frac{-B + \sqrt{B^2 + 4nC}}{2n} \text{ where}$$

$$B = (b+1)\sum_{i=1}^{n}(\alpha_i - (b+1)^{-1})(x_{(i)} - K) \text{ and}$$

$$C = (b+1)\sum_{i=1}^{n}\beta_i (x_{(i)} - K)^2 .$$

Adjusting for the bias we get

$$\hat{\sigma} = \frac{-B + \sqrt{B^2 + 4nC}}{2\sqrt{n(n-1)}} .$$

(3) *Least Squares*

We minimize $E = \sum_{i=1}^{n}(x_i - \mu)^2$.

Setting $\frac{dE}{d\mu} = 0$ we get

$\tilde{\mu} = \bar{x}$.

The LS estimator of σ^2 is

$$\tilde{\sigma}^2 = \min\left[\sum_{i=1}^{n}(x_i - \mu)^2\right] \Big/ (n-1) .$$

Thus, $\tilde{\sigma}^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \Big/ (n-1) = s^2$.

Note that both $\tilde{\mu}$ and $\tilde{\sigma}^2$ are biased. In order to adjust $\tilde{\mu}$ for the bias we use the fact that for the generalized logistic distribution

$E(\bar{x}) = \mu + \big(\psi(b) - \psi(1)\big)\sigma$

and obtain the following unbiased (almost) estimator of μ

$$\tilde{\mu} = \bar{x} - (\psi(b) - \psi(1))\tilde{\sigma} .$$

Now, $\tilde{\sigma}^2$ estimates the population variance which is $\sigma^2(\psi'(b) + \psi'(1))$. Adjusting the above estimator for scale, we get

$$\tilde{\sigma} = \sqrt{\frac{s^2}{(\psi'(b) + \psi'(1))}} .$$

Note: As we saw before, when using a sample of size 100 from the distribution in equation (1.4.1) with b = 1, the MLE of μ and σ obtained have a small bias. However, if the underlying distribution is skew, the iterations converge to values that are far from the true values (μ = 0 and σ = 1). This is shown in the following example.

Example 1.4.2

We generate 100 random numbers from the distribution given in equation (1.4.1) assuming that b = 0.5, μ = 0 and σ = 1. We calculate the ML, MML, and LS estimates for μ and σ. The results are given in Table 1.4.4. Notice that the iterations used to solve the ML equations converged to unrealistic values in this case; MMLE are fin in all respects.

Table 1.4.4 Estimates (GL), n = 100, b = 0.5.

	μ	σ
MLE	-2.891	-1.161
MMLE	-0.086	1.074
LSE	-0.200	0.995

1.4.2 Student's t

Assume that we have a random sample with common p.d.f. given by

$$f(x) = \frac{1}{\sigma\sqrt{r}\ \ B(1/2,\ \ r/2)}\left(1+\frac{1}{r}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-(r+1)/2}. \qquad (1.4.10)$$

Note that the standardized variable $Z = \dfrac{X-\mu}{\sigma}$ in this case has a Student's t distribution with r degrees of freedom.

(1) *Maximum Likelihood*

The likelihood function of $x_1, x_2, \ldots, x_n$ is

$$\mathrm{L} \propto \prod_{i=1}^{n}\left[\frac{1}{\sigma}\left(1+\frac{1}{r}\left(\frac{x_i-\mu}{\sigma}\right)^2\right)^{-(r+1)/2}\right].$$

Taking the logarithm of the likelihood function we get

$$\ln L \propto -n\ln\sigma - \frac{(r+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{z_i^2}{r}\right), \qquad (1.4.11)$$

where $z_i = \dfrac{x_i-\mu}{\sigma}$.

Differentiating equation (1.4.11) with respect to μ and σ we get the following maximum likelihood equations

$$\frac{\partial \ln L}{\partial \mu} = \frac{(r+1)}{r\sigma}\sum_{i=1}^{n}\frac{z_i}{\left(1+\frac{z_i^2}{r}\right)} = 0 \text{ and} \tag{1.4.12}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{(r+1)}{r\sigma}\sum_{i=1}^{n}\frac{z_i^2}{\left(1+\frac{z_i^2}{r}\right)} = 0. \tag{1.4.13}$$

The maximum likelihood method seeks a solution to equations (1.4.12) and (1.4.13). However, again the equations don't have an explicit solution since they are expressions in terms of the intractable function

$$g(z) = \frac{z}{\left(1+\frac{z^2}{r}\right)}.$$

So, the equations must be solved by iteration.

Example 1.4.3

We generated 30 random numbers from the distribution given in equation (1.4.10) assuming r = 4, $\mu = 0$ and σ = 1. The data is given below.

Data set 1:

0.615	0.856	0.211	1.232	0.543	0.572	1.208	0.098	1.639	-1.194
-1.123	0.516	0.092	0.573	-0.139	-3.213	-1.296	1.920	1.396	1.234
0.643	-0.109	-0.255	0.718	-0.952	-0.604	1.411	-5.561	-0.489	0.017

We took equations (1.4.12) and (1.4.13) and solved them by iterations. We got the following solutions to the ML equations:

$\hat{\mu} = 0.291$ and $\hat{\sigma} = 0.918$.

We now include outliers in the data, changing m of the x's to be outliers (m = [0.1n + 0.5]). We generate 30 random numbers again assuming that r = 4, μ = 0 and σ = 1. The data generated is as follows:

Data set 2:

2.461	0.543	0.211	0.856	0.572	1.232	4.832	-1.123	1.639	0.098
0.516	-1.194	0.370	-1.296	-0.139	0.573	1.920	-3.213	1.396	-0.255
0.643	1.234	0.718	-0.109	-0.952	-0.489	1.411	-0.604	0.017	-5.561

Solving equations (1.4.12) and (1.4.13), again the iterations converge giving the following solutions:

$\hat{\mu} = 0.331$ and $\hat{\sigma} = 1.101$.

Note that the estimate of μ is far from the true value of μ which is 0 and $\hat{\sigma}$ = 1.101 is not a good estimate of the true value of σ which is in this case 1.58. Thus, in the presence of outliers, iterations have converged to wrong values.

(2) *Modified Maximum Likelihood*

We express the likelihood equations (1.4.12)-(1.4.13) in terms of the order statistics as follows

$$\frac{\partial \ln L}{\partial \mu} = \frac{(r+1)}{r\sigma} \sum_{i=1}^{n} g(z_{(i)}) = 0 \quad \text{and} \tag{1.4.14}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{(r+1)}{r\sigma} \sum_{i=1}^{n} z_{(i)} g\left(z_{(i)}\right) = 0, \tag{1.4.15}$$

where $z_{(i)} = \dfrac{x_{(i)} - \mu}{\sigma}$. Now, we approximate the function

$$g(z_{(i)}) = \frac{z_{(i)}}{\left(1 + \dfrac{z_{(i)}^2}{r}\right)}$$

by using the first two terms of its Taylor expansion around $t_{(i)} = E(Z_{(i)})$ as follows:

$$g(z_{(i)}) \cong \alpha_i + \beta_i z_{(i)} .$$

To determine the values of alphas and betas we write:

$$g(z_{(i)}) \cong g(t_{(i)}) + (z_{(i)} - t_{(i)}) g'(t_{(i)}) \text{, or}$$

$$g(z_{(i)}) \cong \frac{t_{(i)}}{\left(1 + \dfrac{t_{(i)}^2}{r}\right)} + (z_{(i)} - t_{(i)}) \frac{\left(1 - (t_{(i)}^2 / r)\right)}{\left(1 + \dfrac{t_{(i)}^2}{r}\right)^2} .$$

We get the following equations for alphas and betas,

$$\alpha_i = \frac{2t_{(i)}^3 / r}{\left(1 + \dfrac{t_{(i)}^2}{r}\right)^2} \text{ and } \beta_i = \frac{1 - (t_{(i)}^2 / r)}{\left(1 + \dfrac{t_{(i)}^2}{r}\right)^2},$$

where $t_{(i)}$'s can be obtained from the following equation:

$$\frac{1}{\sqrt{r}\ B\left(1/2,\ r/2\right)} \int_{-\infty}^{t_{(i)}} \left(1 + \frac{t^2}{r}\right)^{-(r+1)/2} dt = \frac{i}{(n+1)}, \quad 1 \le i \le n .$$

Replacing the function g by its linear approximation in equations (1.4.14) and (1.4.15), we get the following modified maximum likelihood equations:

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{(r+1)}{r\sigma}\sum_{i=1}^{n}(\alpha_i + \beta_i z_{(i)}) = 0 \qquad (1.4.16)$$

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{(r+1)}{r\sigma}\sum_{i=1}^{n} z_i(\alpha_i + \beta_i z_{(i)}) = 0. \qquad (1.4.17)$$

Solving equation (1.4.16), we get the following MML estimator of μ:

$$\hat{\mu} = \frac{1}{m}\sum_{i=1}^{n}\beta_i x_{(i)} \quad (m = \sum_{i=1}^{n}\beta_i).$$

Note that, for symmetric distributions, $\sum_{i=1}^{n}\alpha_i = 0.$

To get the MML estimator of σ, we solve equation (1.4.17). The equation can be written as

$$-n\sigma^2 + \frac{(r+1)}{r}\left[\sigma\sum_{i=1}^{n}\alpha_i(x_{(i)} - \mu) + \sum_{i=1}^{n}\beta_i(x_{(i)} - \mu)^2\right] = 0.$$

Replacing μ by $\hat{\mu}$ and noting that $\sum_{i=1}^{n}\alpha_i = 0$, we get

$$-n\hat{\sigma}^2 + \frac{(r+1)\hat{\sigma}}{r}\sum_{i=1}^{n}\alpha_i x_{(i)} + \sum_{i=1}^{n}\beta_i(x_{(i)} - \hat{\mu})^2 = 0,$$ or equivalently

$$n\hat{\sigma}^2 - B\hat{\sigma} - C = 0,$$

which is a quadratic equation with the following admissible solution

$$\hat{\sigma} = \frac{B + \sqrt{B^2 + 4nC}}{2n}$$ where

$$B = \frac{(r+1)}{r}\sum_{i=1}^{n}\alpha_i x_{(i)} \text{ and } C = \frac{(r+1)}{r}\sum_{i=1}^{n}\beta_i (x_{(i)} - \hat{\mu})^2 .$$

Adjusting for the bias, $\hat{\sigma}$ becomes

$$\hat{\sigma} = \frac{B + \sqrt{B^2 + 4nC}}{2\sqrt{n(n-1)}} .$$

Note that if $\beta_i < 0$ for some i, then $\hat{\sigma}$ ceases to be real and positive. In that case we replace the values of alphas and betas above by

$$\alpha_i^* = \frac{t_{(i)}^3 / r}{\left(1 + \frac{t_{(i)}^2}{r}\right)^2} \text{ and } \beta_i^* = \frac{1}{\left(1 + \frac{t_{(i)}^2}{r}\right)^2} .$$

Using the α_i^*'s and β_i^*'s above does not alter the asymptotic properties of the MML estimators (see Tiku et al., 2000). This can be seen by writing

$$g(z_{(i)}) \cong \alpha_i + \beta_i z_{(i)}$$

$$= \frac{2t_{(i)}^3 / r}{\left(1 + \frac{t_{(i)}^2}{r}\right)^2} + \frac{\left(1 - (t_{(i)}^2 / r)\right)}{\left(1 + \frac{t_{(i)}^2}{r}\right)^2} z_{(i)}$$

$$= \frac{2t_{(i)}^3 / r}{\left(1 + \frac{t_{(i)}^2}{r}\right)^2} - \frac{(t_{(i)}^2 / r)}{\left(1 + \frac{t_{(i)}^2}{r}\right)^2} z_{(i)} + \frac{1}{\left(1 + \frac{t_{(i)}^2}{r}\right)^2} z_{(i)}$$

and since $z_{(i)}$ converges to its expected value ($t_{(i)}$) for large n, $g(z_{(i)})$ can be written as

$$g(z_{(i)}) \cong \frac{2t_{(i)}^3 / r}{\left(1+\frac{t_{(i)}^2}{r}\right)^2} - \frac{t_{(i)}^3 / r}{\left(1+\frac{t_{(i)}^2}{r}\right)^2} + \frac{1}{\left(1+\frac{t_{(i)}^2}{r}\right)^2} \quad z_{(i)}$$

$$\cong \frac{t_{(i)}^3 / r}{\left(1+\frac{t_{(i)}^2}{r}\right)^2} + \frac{1}{\left(1+\frac{t_{(i)}^2}{r}\right)^2} \quad z_{(i)} \quad = \quad \alpha_i^* + \beta_i^* z_{(i)}.$$

However, β_1 is less than zero only if r is small (< 5) and n is large (> 50). Therefore, β_i^* and α_i^* need not be used that often.

(3) <u>*Least Squares*</u>

Minimizing E as before we get

$\tilde{\mu} = \bar{x}$

and $\tilde{\sigma}^2 = s^2$.

Note that in equation (1.4.1) if we make the transformation $z = \frac{x-\mu}{\sigma}$ then Z has a Student's t distribution with r degrees of freedom. Which means that the population variance is not equal to σ^2.

$Var(X_i) = \sigma^2 Var(Z_i) = \frac{r}{(r-2)}\sigma^2$.

Thus, the sample mean is an unbiased estimator of μ, however, s^2 is not unbiased for σ^2. It must be adjusted for the bias. So we let

$\tilde{\sigma} = \sqrt{\frac{(r-2)s^2}{r}}$.

We compare the estimates of $\hat{\mu}$ and $\hat{\sigma}$ obtained from the above three methods using data set 1 and data set 2. Table 1.4.5 gives the ML, MML and LS estimates of $\hat{\mu}$ and $\hat{\sigma}$ that are obtained from data set 1.

Table 1.4.5 Estimates from data set 1.

	μ	σ
MLE	0.291	0.918
MMLE	0.222	1.114
LSE	0.019	1.501

Table 1.4.6 gives the ML, MML and LS estimates of $\hat{\mu}$ and $\hat{\sigma}$ that are obtained from data set 2.

Table 1.4.6 Estimates from data set 2.

	μ	σ
MLE	0.331	1.101
MMLE	0.304	1.311
LSE	0.210	1.778

1.4.3 Short Tail Symmetric

Assume we have a random sample that comes from the following family of distributions

$$f(x) \propto \left[1+\frac{\lambda}{2r}\left(\frac{x-\mu}{\sigma}\right)^2\right]^r \frac{1}{\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]. \tag{1.4.18}$$

The family includes unimodal (for $\lambda \le 1$) and bimodal (for $\lambda > 1$) distributions.

(1) <u>*Maximum Likelihood*</u>

The log-likelihood function is

$$\ln L \propto -n\ln\sigma - \frac{1}{2}\sum_{i=1}^{n} z_i^2 + r\sum_{i=1}^{n}\ln\left[1+\frac{\lambda}{2r}z_i^2\right]. \tag{1.4.19}$$

Differentiating equation (1.4.19) with respect to μ and σ we get the following maximum likelihood equations

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{n} z_i - \frac{\lambda}{\sigma}\sum_{i=1}^{n}\frac{z_i}{\left(1+\frac{\lambda}{2r}z_i^2\right)} = 0 \text{ and} \tag{1.4.20}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_i^2 - \frac{\lambda}{\sigma}\sum_{i=1}^{n}\frac{z_i^2}{\left(1+\frac{\lambda}{2r}z_i^2\right)} = 0. \tag{1.4.21}$$

Equations (1.4.20) and (1.4.21) are expressions in terms of the intractable function

$$g(z) = \frac{z}{\left[1+\frac{\lambda}{2r}z^2\right]}.$$

So, once again they don't have an explicit solution and must be solved by iteration.

(2) *Modified Maximum Likelihood*

We express the likelihood equations in terms of order statistics. We then approximate the function

$$g(z_{(i)}) = \frac{z_{(i)}}{\left[1 + \frac{\lambda}{2r} z_{(i)}^2\right]}$$

by the first two terms of its Taylor expansion around $t_{(i)} = E(Z_{(i)})$ as follows

$g(z_{(i)}) \cong \alpha_i + \beta_i z_{(i)}$.

We get the following values for alphas and betas:

For $\lambda \leq 1$

$$\alpha_i = \frac{\lambda t_{(i)}^3 / r}{\left(1 + \frac{\lambda}{2r} t_{(i)}^2\right)^2} \text{ and } \beta_i = \frac{1 - (\lambda t_{(i)}^2 / 2r)}{\left(1 + \frac{\lambda}{2r} t_{(i)}^2\right)^2},$$

and for $\lambda > 1$

$$\alpha_i = \frac{(\lambda t_{(i)}^3 / r) + (1 - 1/\lambda) t_{(i)}}{\left(1 + \frac{\lambda}{2r} t_{(i)}^2\right)^2} \text{ and } \beta_i = \frac{(1/\lambda) - (\lambda t_{(i)}^2 / 2r)}{\left(1 + \frac{\lambda}{2r} t_{(i)}^2\right)^2},$$

where the $t_{(i)}$'s can be obtained from the following equation

$$\frac{c}{\sqrt{2\pi}}\int_{-\infty}^{t_{(i)}}\left(1+\frac{\lambda}{2r}z^2\right)^r e^{-z^2/2}dz=\frac{i}{(n+1)},\quad 1\le i\le n;$$

$$c=1/\left[\sum_{k=0}^{2}\binom{2}{k}\left(\frac{\lambda}{2r}\right)^k\frac{(2k)!}{2^k k!}\right].$$

Writing the equations (1.4.20) and (1.4.21) in terms of order statistics and then replacing the function g by its linear approximation, we get the following modified maximum likelihood equations:

$$\frac{\partial \ln L^*}{\partial \mu}=\frac{1}{\sigma}\sum_{i=1}^{n}z_{(i)}-\frac{\lambda}{\sigma}\sum_{i=1}^{n}(\alpha_i+\beta_i z_{(i)})=0 \text{ and} \tag{1.4.22}$$

$$\frac{\partial \ln L^*}{\partial \sigma}=-\frac{n}{\sigma}+\frac{1}{\sigma}\sum_{i=1}^{n}z_{(i)}^2-\frac{\lambda}{\sigma}\sum_{i=1}^{n}z_{(i)}(\alpha_i+\beta_i z_{(i)})=0. \tag{1.4.23}$$

The solutions of equations (1.4.22) and (1.4.23) are the following MML estimators:

$$\hat{\mu}=\frac{1}{m_1}\sum_{i=1}^{n}\beta_{1i}x_{(i)}\ \left(m_1=\sum_{i=1}^{n}\beta_{1i}\ ,\ \beta_{1i}=1-\lambda\beta_i\right) \text{ and}$$

$$\hat{\sigma}=\frac{-\lambda B+\sqrt{(\lambda B)^2+4nC}}{2n}, \text{ where}$$

$$B=\sum_{i=1}^{n}\alpha_i x_{(i)} \text{ and } C=\sum_{i=1}^{n}\beta_{1i}(x_{(i)}-\hat{\mu})^2 .$$

Adjusting for the bias we get

$$\hat{\sigma}=\frac{-\lambda B+\sqrt{(\lambda B)^2+4nC}}{2\sqrt{n(n-1)}} .$$

(3) *Least Squares*

Minimizing E we get

$\tilde{\mu} = \bar{x}$ and $\tilde{\sigma}^2 = s^2$;

$\tilde{\mu}$ is unbiased for μ. However,

$Var(X_i) = \sigma^2 Var(Z_i) = \sigma^2 W$, where

$$W = Var(Z_i) = \left[\sum_{j=0}^{r} \binom{r}{j} \left(\frac{\lambda}{2r}\right)^j \frac{(2(j+1))!}{2^{j+1}(j+1)!}\right] \Bigg/ \left[\sum_{j=0}^{r} \binom{r}{j} \left(\frac{\lambda}{2r}\right)^j \frac{(2j)!}{2^j j!}\right].$$

Therefore, we must adjust the estimator of σ accordingly. The adjusted LSE of σ is

$$\tilde{\sigma} = \sqrt{\frac{s^2}{W}}.$$

Remark: The modified maximum likelihood methodology readily extends to censored samples and to other areas, e.g., experimental design, time series, regression, etc. (see for example Akkaya and Tiku, 2005, Bhattacharyya, 1985, Islam and Tiku, 2004). We will utilize this method for estimating parameters in bivariate distributions.

CHAPTER 2

NON-NORMAL BIVARIATE DISTRIBUTIONS

Introduction

In Chapter 1 we considered a few univariate distributions and illustrated how each method finds estimators of the population parameters. Here we will consider bivariate distributions. Vaughan and Tiku (2000), Sazak et al.(2006) and Tiku et al. (2007) considered the situation when the bivariate random variable (X,Y) is such that Y depends on X (explicitly or implicitly) and not so much the other way around. Specifically, E(Y/X = x) is a linear function of x and the conditional variance of Y given X = x is constant or a constant multiple of $w(x)$ (a positive function of x). They used the modified maximum likelihood method to find estimators of the population parameters under any location-scale non-normal bivariate distribution. Here we consider the situation when we have two independent sets of bivariate data. Thus, we extend the work from single-sample to two-sample situations.

We consider two bivariate random vectors (X, Y) and (U, V) when E(Y/X = x) is a linear function of x and E(V/U = u) is a linear function of u. As for the conditional variances, we consider two situations, when Var(Y/X = x) is constant, and when it is a constant multiple of a positive function of x (similarly for Var(V/U = u)). Specifically, we consider two distinctive distributions: the marginal distribution of X and U and the conditional distribution of Y and V are both Generalized Logistic, and the marginal and conditional both belong to the Student's t family. It may be noted that even though we consider only the above bivariate distributions, this work can be extended to any other bivariate distribution. We use the method of modified maximum likelihood to find estimators of various parameters in each distribution.

2.1 Models

Suppose we have two random vectors, (X, Y) and (U, V). Suppose that Y depends on X, and not so much the other way around. Similarly, suppose that V depends on U.

Therefore, assume we have the following models:

Model 1: $E(Y/X=x)=\mu_y+\rho_{yx}\frac{\sigma_y}{\sigma_x}(x-\mu_x)=\mu_{y/x}+\theta_{yx}x$. (2.1.1)

Model 2: $E(V/U=u)=\mu_v+\rho_{vu}\frac{\sigma_v}{\sigma_u}(u-\mu_u)=\mu_{v/u}+\theta_{vu}u$. (2.1.2)

$\mu_{y/x}=\mu_y-\theta_{yx}\mu_x$, $\theta_{yx}=\rho_{yx}\frac{\sigma_y}{\sigma_x}$,

$\mu_{v/u}=\mu_v-\theta_{vu}\mu_u$ and $\theta_{vu}=\rho_{vu}\frac{\sigma_v}{\sigma_u}$;

Y, X, V and U are all stochastic variables. We take a random sample of size n_1 from model 1 and a sample of size n_2 from model 2. We are interested in estimating all the above parameters.

At first we will assume equal variances and covariances and also equal sample sizes. Thus, we assume that:

$\rho_{yx} = \rho_{vu} = \rho$,

$\sigma_y = \sigma_v = \sigma_2$,

$\sigma_x = \sigma_u = \sigma_1$, and

$n_1 = n_2 = n$.

In this case, the above models reduce to:

Model 1: $E(Y / X = x) = \mu_y + \rho \dfrac{\sigma_2}{\sigma_1}(x - \mu_x) = \mu_{y/x} + \theta\, x$. (2.1.3)

Model 2: $E(V / U = u) = \mu_v + \rho \dfrac{\sigma_2}{\sigma_1}(u - \mu_u) = \mu_{v/u} + \theta\, u$. (2.1.4)

2.2 Normal Marginal and Conditional

If the joint distribution of X and Y is bivariate normal, and the joint distribution of U and V is also bivariate normal, then the least square estimators obtained from the two random samples (x_i, y_i) and (u_i, v_i) $(1 \le i \le n)$ are exactly the same as the maximum likelihood estimators. That is, if X is distributed as normal N(μ_x, σ_1^2), U~N(μ_u, σ_1^2), e_1~N($0, \sigma_{2.1}$) and e_2~N($0, \sigma_{2.1}$), where $e_1 = Y - \mu_{y/x} - \theta\, x$ and $e_2 = V - \mu_{v/u} - \theta\, u$, then the LSE are the same as the MLE and are, therefore, fully efficient. However, if one or more distributions are non-normal, the LS estimators loose their efficiency and also develop bias. We can easily adjust the LSE for the bias, however, they loose their efficiency particularly for large n and that is a very undesirable property in an estimator. The LSE are also not robust to deviations from an assumed distribution and not robust to the presence of outliers in the data and other anomalies. We use the method of modified likelihood to find alternative estimators of the parameters that are efficient and robust and have many other desirable properties (Vaughan, 2002).

2.3 Non-normal Marginal and Conditional

As we mentioned before, if the conditional and marginal distributions are not normal, we often face situations where the maximum likelihood equations have no explicit solutions. Also, the least square estimators loose their efficiency under non-normality. Thus, in case of non-normality, we use the modified maximum likelihood method. Here, we consider two situations: When the marginal and conditional distributions are both Generalized logistic, and when they belong to the Student's t family. It may be noted, however, that the modified maximum likelihood method can be extended to any location-scale parameter family.

2.4 Generalized Logistic Distribution

At first, let us assume that the marginal and conditional distributions are both generalized logistic. We assume the following:

(a) The marginal distribution of X is Generalized logistic with shape parameter b_x (>0), i.e.,

$$f_x(x) = \frac{b_x}{\sigma_1} \frac{\exp\left(-\frac{x-\mu_x}{\sigma_1}\right)}{\left[1+\exp\left(-\frac{x-\mu_x}{\sigma_1}\right)\right]^{b_x+1}} . \qquad (2.4.1)$$

(b) The conditional distribution of the error e_1 (the distribution of Y given X = *x*) is Generalized logistic with shape parameter b_y (>0), i.e.,

$$f_1(e_1) = \frac{b_y}{\sigma_{2.1}} \frac{\exp\left(-\frac{e_1}{\sigma_{2,1}}\right)}{\left[1+\exp\left(-\frac{e_1}{\sigma_{2.1}}\right)\right]^{b_y+1}}. \tag{2.4.2}$$

(c) The marginal distribution of U is Generalized logistic with shape parameter b_u (>0), i.e.,

$$f_u(u) = \frac{b_u}{\sigma_1} \frac{\exp\left(-\frac{u-\mu_u}{\sigma_1}\right)}{\left[1+\exp\left(-\frac{u-\mu_u}{\sigma_1}\right)\right]^{b_u+1}}. \tag{2.4.3}$$

(d) The conditional distribution of the error e_2 (the distribution of V given U = *u*) is Generalized logistic with shape parameter b_v (>0), i.e.,

$$f_2(e_2) = \frac{b_v}{\sigma_{2.1}} \frac{\exp\left(-\frac{e_2}{\sigma_{2.1}}\right)}{\left[1+\exp\left(-\frac{e_2}{\sigma_{2.1}}\right)\right]^{b_v+1}}. \tag{2.4.4}$$

2.4.1 Modified Maximum likelihood

Given two random samples:
$(x_i, y_i), 1 \le i \le n$, from the first model and $(u_i, v_i), 1 \le i \le n$, from the second model.

The likelihood function is:

$\mathrm{L} = L_x L_{y/x} L_u L_{v/u} = L_x L_{e_1} L_u L_{e_2}$.

Thus we have,

$$\mathrm{L} \propto \prod_{i=1}^{n} \left[\frac{1}{\sigma_1} \frac{\exp\left(-\frac{x_i-\mu_x}{\sigma_1}\right)}{\left[1+\exp\left(-\frac{x_i-\mu_x}{\sigma_1}\right)\right]^{b_x+1}} \; \frac{1}{\sigma_{2.1}} \frac{\exp\left(-\frac{e_{1i}}{\sigma_{2.1}}\right)}{\left[1+\exp\left(-\frac{e_{1i}}{\sigma_{2.1}}\right)\right]^{b_y+1}} \right.$$

$$\left. \frac{1}{\sigma_1} \frac{\exp\left(-\frac{u_i-\mu_u}{\sigma_1}\right)}{\left[1+\exp\left(-\frac{u_i-\mu_u}{\sigma_1}\right)\right]^{b_u+1}} \; \frac{1}{\sigma_{2.1}} \frac{\exp\left(-\frac{e_{2i}}{\sigma_{2.1}}\right)}{\left[1+\exp\left(-\frac{e_{2i}}{\sigma_{2.1}}\right)\right]^{b_v+1}} \right]. \qquad (2.4.5)$$

Let,

$Z_1 = \frac{X-\mu_x}{\sigma_1}, Z_2 = \frac{e_1}{\sigma_{2.1}} = \frac{Y-\mu_{y/x}-\theta\, x}{\sigma_{2.1}}, W_1 = \frac{U-\mu_u}{\sigma_1}$ and

$W_2 = \frac{e_2}{\sigma_{2.1}} = \frac{V-\mu_{v/u}-\theta\, u}{\sigma_{2.1}}$.

We can now express the log likelihood in terms of $z_{1i},\ z_{2i}, w_{1i},\ w_{2i}$ as follows:

$$\ln L \propto -2n\ln\sigma_1 - 2n\ln\sigma_{2.1} - \sum_{i=1}^{n} z_{1i} - (b_x+1)\sum_{i=1}^{n}\ln(1+e^{-z_{1i}})$$

$$-\sum_{i=1}^{n} z_{2i} - (b_y+1)\sum_{i=1}^{n}\ln(1+e^{-z_{2i}}) - \sum_{i=1}^{n} w_{1i} - (b_u+1)\sum_{i=1}^{n}\ln(1+e^{-w_{1i}})$$

$$-\sum_{i=1}^{n} w_{2i} - (b_v+1)\sum_{i=1}^{n}\ln(1+e^{-w_{2i}}), \qquad (2.4.6)$$

where

$$z_{1i} = \frac{x_i - \mu_x}{\sigma_1}, z_{2i} = \frac{e_{1i}}{\sigma_{2.1}} = \frac{y_i - \mu_{y/x} - \theta\, x_i}{\sigma_{2.1}}, w_{1i} = \frac{u_i - \mu_u}{\sigma_1},$$

and $w_{2i} = \frac{e_{2i}}{\sigma_{2.1}} = \frac{v_i - \mu_{v/u} - \theta\, u_i}{\sigma_{2.1}}$.

We differentiate equation (2.4.6) to find the following likelihood equations which can be used for estimating μ_x, μ_u, $\mu_{y/x}$, $\mu_{v/u}$, σ_1, $\sigma_{2.1}$, θ:

$$\frac{\partial \ln L}{\partial \mu_x} = \frac{n}{\sigma_1} - \frac{(b_x+1)}{\sigma_1}\sum_{i=1}^{n}\frac{e^{-z_{1i}}}{(1+e^{-z_{1i}})} - \frac{n\theta}{\sigma_{2.1}} + \frac{(b_y+1)\theta}{\sigma_{2.1}}\sum_{i=1}^{n}\frac{e^{-z_{2i}}}{(1+e^{-z_{2i}})} = 0 \quad (2.4.7)$$

$$\frac{\partial \ln L}{\partial \mu_u} = \frac{n}{\sigma_1} - \frac{(b_u+1)}{\sigma_1}\sum_{i=1}^{n}\frac{e^{-w_{1i}}}{(1+e^{-w_{1i}})} - \frac{n\theta}{\sigma_{2.1}} + \frac{(b_v+1)\theta}{\sigma_{2.1}}\sum_{i=1}^{n}\frac{e^{-w_{2i}}}{(1+e^{-w_{2i}})} = 0 \quad (2.4.8)$$

$$\frac{\partial \ln L}{\partial \mu_{y/x}} = \frac{n}{\sigma_{2.1}} - \frac{(b_y+1)}{\sigma_{2.1}}\sum_{i=1}^{n}\frac{e^{-z_{2i}}}{(1+e^{-z_{2i}})} = 0 \quad (2.4.9)$$

$$\frac{\partial \ln L}{\partial \mu_{v/u}} = \frac{n}{\sigma_{2.1}} - \frac{(b_v+1)}{\sigma_{2.1}}\sum_{i=1}^{n}\frac{e^{-w_{2i}}}{(1+e^{-w_{2i}})} = 0 \quad (2.4.10)$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{2n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_{1i} - \frac{(b_x+1)}{\sigma_1}\sum_{i=1}^{n}\frac{z_{1i}e^{-z_{1i}}}{(1+e^{-z_{1i}})} - \frac{\theta}{\sigma_{2.1}}\sum_{i=1}^{n} z_{1i}$$
$$+ \frac{(b_y+1)\theta}{\sigma_{2.1}}\sum_{i=1}^{n}\frac{z_{1i}e^{-z_{2i}}}{(1+e^{-z_{2i}})} + \frac{1}{\sigma_1}\sum_{i=1}^{n} w_{1i} - \frac{(b_u+1)}{\sigma_1}\sum_{i=1}^{n}\frac{w_{1i}e^{-w_{1i}}}{(1+e^{-w_{1i}})}$$
$$- \frac{\theta}{\sigma_{2.1}}\sum_{i=1}^{n} w_{1i} + \frac{(b_v+1)\theta}{\sigma_{2.1}}\sum_{i=1}^{n}\frac{w_{1i}e^{-w_{2i}}}{(1+e^{-w_{2i}})} = 0 \quad (2.4.11)$$

$$\frac{\partial \ln L}{\partial \sigma_{2.1}} = -\frac{2n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}}\sum_{i=1}^{n} z_{2i} - \frac{(b_y+1)}{\sigma_{2.1}}\sum_{i=1}^{n}\frac{z_{2i}e^{-z_{2i}}}{(1+e^{-z_{2i}})}$$

$$+\frac{1}{\sigma_{2.1}}\sum_{i=1}^{n} w_{2i} - \frac{(b_v+1)}{\sigma_{2.1}}\sum_{i=1}^{n}\frac{w_{2i}e^{-w_{2i}}}{(1+e^{-w_{2i}})} = 0 \qquad (2.4.12)$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} z_{1i} - \frac{(b_y+1)\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n}\frac{z_{1i}e^{-z_{2i}}}{(1+e^{-z_{2i}})}$$

$$+\frac{\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} w_{1i} - \frac{(b_y+1)\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n}\frac{w_{1i}e^{-w_{2i}}}{(1+e^{-w_{2i}})} = 0. \qquad (2.4.13)$$

The above likelihood equations are expressions in terms of the intractable functions:

$$g_1(z_{1i}) = \frac{e^{-z_{1i}}}{1+e^{-z_{1i}}},\ g_1(z_{2i}) = \frac{e^{-z_{2i}}}{1+e^{-z_{2i}}},$$

$$h_1(w_{1i}) = \frac{e^{-w_{1i}}}{1+e^{-w_{1i}}} \text{ and } h_2(w_{2i}) = \frac{e^{-w_{2i}}}{1+e^{-w_{2i}}},$$

and have no explicit solutions. Solving them by iteration might be problematic due to many reasons such as: multiple roots, convergence to wrong values and nonconvergence, as illustrated in chapter 1.

We use the method of modified maximum likelihood estimation which originated with Tiku (1967, 1989) and Tiku & Suresh (1992) as said earlier, to find solutions of the above equations. First we order x_i's, e_{1i}'s, u_i's and e_{2i}'s in ascending order of magnitude as follows:

$$x_{(1)} \le x_{(2)} \le ... \le x_{(n)},$$

$$e_{1(1)} \le e_{1(2)} \le ... \le e_{1(n)},$$

$$u_{(1)} \le u_{(2)} \le ... \le u_{(n)} \text{ and}$$

$$e_{2(1)} \le e_{2(2)} \le ... \le e_{2(n)}.$$

Note that $z_{1(i)}$ has the same order as $x_{(i)}$ since $z_{1(i)} = \dfrac{x_{(i)} - \mu_x}{\sigma_1}$ and μ_x is a constant and σ_1 is positive. For similar reasons, $z_{2(i)}$ has the same order as $e_{1(i)}$; $w_{1(i)}$ has the same order as $u_{(i)}$, and $w_{2(i)}$ has the same order as $e_{2(i)}$.

Now if $e_{1(i)} = y_{[i]} - \mu_{y/x} - \theta\, x_{[i]}$ then we say $(x_{[i]}, y_{[i]})$ are the concomitants of $e_{1(i)}$. Also, if $e_{2(i)} = v_{[i]} - \mu_{v/u} - \theta\, u_{[i]}$ then we say $(u_{[i]}, v_{[i]})$ are the concomitants of $e_{2(i)}$.

We now express our likelihood in terms of the order statistics as follows:

$$\ln L \propto -2n\ln\sigma_1 - 2n\ln\sigma_{2.1} - \sum_{i=1}^{n} z_{1(i)} - (b_x+1)\sum_{i=1}^{n}\ln(1+e^{-z_{1(i)}})$$
$$-\sum_{i=1}^{n} z_{2(i)} - (b_y+1)\sum_{i=1}^{n}\ln(1+e^{-z_{2(i)}}) - \sum_{i=1}^{n} w_{1(i)}$$
$$-(b_u+1)\sum_{i=1}^{n}\ln(1+e^{-w_{1(i)}}) - \sum_{i=1}^{n} w_{2(i)} - (b_v+1)\sum_{i=1}^{n}\ln(1+e^{-w_{2(i)}})\,. \tag{2.4.14}$$

Note that this does not result in any change in the numerical values of the likelihood since the above are all complete sums.

The maximum likelihood equations become:

$$\frac{\partial \ln L}{\partial \mu_x} = \frac{n}{\sigma_1} - \frac{(b_x+1)}{\sigma_1}\sum_{i=1}^{n} g_1(z_{1(i)}) - \frac{n\theta}{\sigma_{2.1}} + \frac{(b_y+1)\theta}{\sigma_{2.1}}\sum_{i=1}^{n} g_2(z_{2(i)}) = 0 \tag{2.4.15}$$

$$\frac{\partial \ln L}{\partial \mu_u} = \frac{n}{\sigma_1} - \frac{(b_u+1)}{\sigma_1}\sum_{i=1}^{n} h_1(w_{1(i)}) - \frac{n\theta}{\sigma_{2.1}} + \frac{(b_v+1)\theta}{\sigma_{2.1}}\sum_{i=1}^{n} h_2(w_{2(i)}) = 0 \qquad (2.4.16)$$

$$\frac{\partial \ln L}{\partial \mu_{y/x}} = \frac{n}{\sigma_{2.1}} - \frac{(b_y+1)}{\sigma_{2.1}}\sum_{i=1}^{n} g_2(z_{2(i)}) = 0 \qquad (2.4.17)$$

$$\frac{\partial \ln L}{\partial \mu_{v/u}} = \frac{n}{\sigma_{2.1}} - \frac{(b_v+1)}{\sigma_{2.1}}\sum_{i=1}^{n} h_2(w_{2(i)}) = 0 \qquad (2.4.18)$$

$$\begin{aligned}\frac{\partial \ln L}{\partial \sigma_1} = &-\frac{2n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_{1(i)} - \frac{(b_x+1)}{\sigma_1}\sum_{i=1}^{n} z_{1(i)} g_1(z_{1(i)}) - \frac{\theta}{\sigma_{2.1}}\sum_{i=1}^{n} z_{1(i)} \\ &+ \frac{(b_y+1)\theta}{\sigma_{2.1}}\sum_{i=1}^{n} z_{1(i)} g_2(z_{2(i)}) + \frac{1}{\sigma_1}\sum_{i=1}^{n} w_{1(i)} - \frac{(b_u+1)}{\sigma_1}\sum_{i=1}^{n} w_{1(i)} h_1(w_{1(i)}) \\ &- \frac{\theta}{\sigma_{2.1}}\sum_{i=1}^{n} w_{1(i)} + \frac{(b_v+1)\theta}{\sigma_{2.1}}\sum_{i=1}^{n} w_{1(i)} h_2(w_{2(i)}) = 0 \qquad (2.4.19)\end{aligned}$$

$$\begin{aligned}\frac{\partial \ln L}{\partial \sigma_{2.1}} = &-\frac{2n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}}\sum_{i=1}^{n} z_{2(i)} - \frac{(b_y+1)}{\sigma_{2.1}}\sum_{i=1}^{n} z_{2(i)} g_2(z_{2(i)}) + \frac{1}{\sigma_{2.1}}\sum_{i=1}^{n} w_{2(i)} \\ &- \frac{(b_v+1)}{\sigma_{2.1}}\sum_{i=1}^{n} w_{2(i)} h_2(w_{2(i)}) = 0 \qquad (2.4.20)\end{aligned}$$

$$\begin{aligned}\frac{\partial \ln L}{\partial \theta} = &\frac{\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} z_{1(i)} - \frac{(b_y+1)\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} z_{1(i)} g_2(z_{2(i)}) + \frac{\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} w_{1(i)} \\ &- \frac{(b_y+1)\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} w_{1(i)} h_2(w_{2(i)}) = 0. \qquad (2.4.21)\end{aligned}$$

Now, we approximate the functions g_1 and g_2 by the first two terms of their Taylor series expansions around $t_{1(i)} = E(Z_{1(i)})$ and $t_{2(i)} = E(Z_{2(i)})$ respectively:

$$g_1(z_{1(i)}) \cong \alpha_{1i} - \beta_{1i} z_{1(i)},$$

$$g_2(z_{2(i)}) \cong \alpha_{2i} - \beta_{2i} z_{2(i)} .$$

To determine the values of alphas and betas we write:

$$g_1(z_{1(i)}) \cong g_1(t_{1(i)}) + (z_{1(i)} - t_{1(i)}) g_1{}'(t_{1(i)})$$

$$g_1(z_{1(i)}) \cong \frac{e^{-t_{1(i)}}}{(1+e^{-t_{1(i)}})} + (z_{1(i)} - t_{1(i)}) \left(\frac{-e^{-t_{1(i)}}}{(1+e^{-t_{1(i)}})^2} \right).$$

Thus,

$$\beta_{1i} = \frac{e^{-t_{1(i)}}}{(1+e^{-t_{1(i)}})^2} \text{ and } \alpha_{1i} = \frac{e^{-t_{1(i)}}}{(1+e^{-t_{1(i)}})} + \beta_{1i} t_{1(i)} .$$

We do the same for $g_2(z_{2(i)})$:

$$g_2(z_{2(i)}) \cong g_2(t_{2(i)}) + (z_{2(i)} - t_{2(i)}) g_2{}'(t_{2(i)})$$

$$g_2(z_{2(i)}) \cong \frac{e^{-t_{2(i)}}}{(1+e^{-t_{2(i)}})} + (z_{2(i)} - t_{2(i)}) \left(\frac{-e^{-t_{2(i)}}}{(1+e^{-t_{2(i)}})^2} \right).$$

We get

$$\beta_{2i} = \frac{e^{-t_{2(i)}}}{(1+e^{-t_{2(i)}})^2} \text{ and } \alpha_{2i} = \frac{e^{-t_{2(i)}}}{(1+e^{-t_{2(i)}})} + \beta_{2i} t_{2(i)} .$$

Values of $t_{1(i)}$ and $t_{2(i)}$ for n $\leq$ 15 are available in Balakrishnan and Leung (1988). For n ≥ 10, however, we can use the following approximate values of $t_{1(i)}$ and $t_{2(i)}$:

$$t_{1(i)} = -\ln(q_i^{-1/b_x} - 1) ,\; t_{2(i)} = -\ln(q_i^{-1/b_y} - 1) ,\; q_i = i/(n+1) .$$

We also approximate the functions h_1 and h_2 by the first two terms of their Taylor series expansions around $t^*_{1(i)} = E(W_{1(i)})$ and $t^*_{2(i)} = E(W_{2(i)})$ respectively:

$$h_1(w_{1(i)}) \cong \delta_{1i} - \gamma_{1i} w_{1(i)}$$

$$h_2(w_{2(i)}) \cong \delta_{2i} - \gamma_{2i} w_{2(i)}.$$

Using the same procedure as before we get:

$$\gamma_{1i} = \frac{e^{-t^*_{1(i)}}}{(1+e^{-t^*_{1(i)}})^2}, \ \delta_{1i} = \frac{e^{-t^*_{1(i)}}}{(1+e^{-t^*_{1(i)}})} + \gamma_{1i} t^*_{1(i)}$$

$$\gamma_{2i} = \frac{e^{-t^*_{2(i)}}}{(1+e^{-t^*_{2(i)}})^2} \text{ and } \delta_{2i} = \frac{e^{-t^*_{2(i)}}}{(1+e^{-t^*_{2(i)}})} + \gamma_{2i} t^*_{2(i)},$$

where again we can use the following approximate values of $t^*_{1(i)}$ and $t^*_{2(i)}$:

$$t^*_{1(i)} = -\ln(q_i^{-1/b_u} - 1), \ t^*_{2(i)} = -\ln(q_i^{-1/b_v} - 1), \ q_i = i/(n+1).$$

Substituting the linear approximations of the functions g_1, g_2 , h_1 and h_2 in the likelihood equations (2.4.15)-(2.4.21) we obtain the following modified maximum likelihood equations:

$$\frac{\partial \ln L^*}{\partial \mu_x} = \frac{n}{\sigma_1} - \frac{(b_x+1)}{\sigma_1}\sum_{i=1}^{n}(\alpha_{1i} - \beta_{1i} z_{1(i)}) - \frac{n\theta}{\sigma_{2.1}} + \frac{(b_y+1)\theta}{\sigma_{2.1}}\sum_{i=1}^{n}(\alpha_{2i} - \beta_{2i} z_{2(i)})$$

$$=0 \qquad (2.4.22)$$

$$\frac{\partial \ln L^*}{\partial \mu_u} = \frac{n}{\sigma_1} - \frac{(b_u+1)}{\sigma_1}\sum_{i=1}^{n}(\delta_{1i} - \gamma_{1i}w_{1(i)}) - \frac{n\theta}{\sigma_{2.1}} + \frac{(b_v+1)\theta}{\sigma_{2.1}}\sum_{i=1}^{n}(\delta_{2i} - \gamma_{2i}w_{2(i)}) = 0 \tag{2.4.23}$$

$$\frac{\partial \ln L^*}{\partial \mu_{y/x}} = \frac{n}{\sigma_{2.1}} - \frac{(b_y+1)}{\sigma_{2.1}}\sum_{i=1}^{n}(\alpha_{2i} - \beta_{2i}z_{2(i)}) = 0 \tag{2.4.24}$$

$$\frac{\partial \ln L^*}{\partial \mu_{v/u}} = \frac{n}{\sigma_{2.1}} - \frac{(b_v+1)}{\sigma_{2.1}}\sum_{i=1}^{n}(\delta_{2i} - \gamma_{2i}w_{2(i)}) = 0 \tag{2.4.25}$$

$$\begin{aligned}\frac{\partial \ln L^*}{\partial \sigma_1} = &-\frac{2n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_{1(i)} - \frac{(b_x+1)}{\sigma_1}\sum_{i=1}^{n} z_{1(i)}(\alpha_{1i} - \beta_{1i}z_{1(i)}) - \frac{\theta}{\sigma_{2.1}}\sum_{i=1}^{n} z_{1(i)} \\ &+ \frac{(b_y+1)\theta}{\sigma_{2.1}}\sum_{i=1}^{n} z_{1(i)}(\alpha_{2i} - \beta_{2i}z_{2(i)}) + \frac{1}{\sigma_1}\sum_{i=1}^{n} w_{1(i)} - \frac{(b_u+1)}{\sigma_1}\sum_{i=1}^{n} w_{1(i)}(\delta_{1i} - \gamma_{1i}w_{1(i)}) \\ &- \frac{\theta}{\sigma_{2.1}}\sum_{i=1}^{n} w_{1(i)} + \frac{(b_v+1)\theta}{\sigma_{2.1}}\sum_{i=1}^{n} w_{1(i)}(\delta_{2i} - \gamma_{2i}w_{2(i)}) = 0\end{aligned} \tag{2.4.26}$$

$$\begin{aligned}\frac{\partial \ln L^*}{\partial \sigma_{2.1}} = &-\frac{2n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}}\sum_{i=1}^{n} z_{2(i)} - \frac{(b_y+1)}{\sigma_{2.1}}\sum_{i=1}^{n} z_{2(i)}(\alpha_{2i} - \beta_{2i}z_{2(i)}) + \frac{1}{\sigma_{2.1}}\sum_{i=1}^{n} w_{2(i)} \\ &- \frac{(b_v+1)}{\sigma_{2.1}}\sum_{i=1}^{n} w_{2(i)}(\delta_{2i} - \gamma_{2i}w_{2(i)}) = 0\end{aligned} \tag{2.4.27}$$

$$\begin{aligned}\frac{\partial \ln L^*}{\partial \theta} = &\frac{\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} z_{1(i)} - \frac{(b_y+1)\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} z_{1(i)}(\alpha_{2i} - \beta_{2i}z_{2(i)}) + \frac{\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} w_{1(i)} \\ &- \frac{(b_y+1)\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} w_{1(i)}(\delta_{2i} - \gamma_{2i}w_{2(i)}) = 0.\end{aligned} \tag{2.4.28}$$

Unlike the maximum likelihood equations, the above equations have explicit solutions. The solutions are the following MML estimators:

1. $\hat{\mu}_x = K_{11} - D_{11}\hat{\sigma}_1$, where

$$K_{11}=\frac{\sum_{i=1}^{n}\beta_{1i}x_{(i)}}{m_{11}},\ D_{11}=\frac{1}{m_{11}}\sum_{i=1}^{n}\left(\alpha_{1i}-(b_x+1)^{-1}\right) \text{ and } m_{11}=\sum_{i=1}^{n}\beta_{1i}\,.$$

2. $\hat{\mu}_u = K_{21}-D_{21}\hat{\sigma}_1$ where,

$$K_{21}=\frac{\sum_{i=1}^{n}\gamma_{1i}u_{(i)}}{m_{21}},\ D_{21}=\frac{1}{m_{21}}\sum_{i=1}^{n}\left(\delta_{1i}-(b_u+1)^{-1}\right) \text{ and } m_{21}=\sum_{i=1}^{n}\gamma_{1i}\,.$$

3. $\hat{\sigma}_1^* = \dfrac{-B_1+\sqrt{B_1^{\,2}+8nC_1}}{4n}$, where

$$B_1=(b_x+1)\sum_{i=1}^{n}\left[\left(\alpha_{1i}-(b_x+1)^{-1}\right)\left(x_{(i)}-K_{11}\right)\right]$$

$$+(b_u+1)\sum_{i=1}^{n}\left[\left(\delta_{1i}-(b_u+1)^{-1}\right)\left(u_{(i)}-K_{21}\right)\right],$$

$$C_1=(b_x+1)\sum_{i=1}^{n}\beta_{1i}(x_{(i)}-K_{11})^2+(b_u+1)\sum_{i=1}^{n}\gamma_{1i}(u_{(i)}-K_{21})^2\,;$$

adjusting for the bias we get: $\hat{\sigma}_1 = \dfrac{-B_1+\sqrt{B_1^{\,2}+8nC_1}}{4\sqrt{n(n-2)}}$.

4. $\hat{\mu}_{y/x} = \overline{y}_{[.]}-\hat{\theta}\overline{x}_{[.]}-\hat{\sigma}_{2.1}\Delta_1/m_{12}$, where

$$\Delta_1=\sum_{i=1}^{n}\Delta_{1i}\,,\ \ \Delta_{1i}=(\alpha_{2i}-(b_y+1)^{-1})\,,$$

$$\overline{y}_{[.]}=\frac{1}{m_{12}}\sum\beta_{2i}y_{[i]}\,,\ \overline{x}_{[.]}=\frac{1}{m_{12}}\sum\beta_{2i}x_{[i]}\,;\ m_{12}=\sum_{i=1}^{n}\beta_{2i}\,.$$

5. $\hat{\mu}_{v/u} = \overline{v}_{[.]}-\hat{\theta}\overline{u}_{[.]}-\hat{\sigma}_{2.1}\Delta_2/m_{22}$, where

$$\Delta_2 = \sum_{i=1}^{n} \Delta_{2i} \ , \ \ \Delta_{2i} = (\delta_{2i} - (b_v + 1)^{-1}) \ ,$$

$$\overline{v}_{[.]} = \frac{1}{m_{22}} \sum_{i=1}^{n} \gamma_{2i} v_{[i]} \ , \ \overline{u}_{[.]} = \frac{1}{m_{22}} \sum_{i=1}^{n} \gamma_{2i} u_{[i]} \ ; \ m_{22} = \sum_{i=1}^{n} \gamma_{2i} \ .$$

6. $\hat{\theta} = K - D \ \hat{\sigma}_{2.1}$, where

$$K = \frac{1}{S}\left[(b_y + 1) \sum_{i=1}^{n} \beta_{2i} (x_{[i]} - \overline{x}_{[.]}) y_{(i)} + (b_v + 1) \sum_{i=1}^{n} \gamma_{2i} (u_{[i]} - \overline{u}_{[.]}) v_{(i)} \right],$$

$$D = \frac{1}{S}\left[(b_y + 1) \sum_{i=1}^{n} \Delta_{1i} (x_{[i]} - \overline{x}_{[.]}) + (b_v + 1) \sum_{i=1}^{n} \Delta_{2i} (u_{[i]} - \overline{u}_{[.]}) \right],$$

$$S = (b_y + 1) \sum_{i=1}^{n} \beta_{2i} (x_{[i]} - \overline{x}_{[.]})^2 + (b_v + 1) \sum_{i=1}^{n} \gamma_{2i} (u_{[i]} - \overline{u}_{[.]})^2 \ .$$

7. $\hat{\sigma}^*_{2.1} = \dfrac{-B + \sqrt{B^2 + 8nC}}{4n}$, where

$$B = (b_y + 1) \sum_{i=1}^{n} \Delta_{1i} (y_{[i]} - \overline{y}_{[.]} - K(x_{[i]} - \overline{x}_{[.]}))$$

$$+ (b_v + 1) \sum_{i=1}^{n} \Delta_{2i} (v_{[i]} - \overline{v}_{[.]} - K(u_{[i]} - \overline{u}_{[.]})) \ ,$$

$$C = (b_y + 1) \sum_{i=1}^{n} \beta_{2i} (y_{[i]} - \overline{y}_{[.]} - K(x_{[i]} - \overline{x}_{[.]}))^2$$

$$+ (b_v + 1) \sum_{i=1}^{n} \gamma_{2i} (v_{[i]} - \overline{v}_{[.]} - K(u_{[i]} - \overline{u}_{[.]}))^2 \ .$$

Adjusting for the bias we get: $\hat{\sigma}_{2.1} = \dfrac{-B + \sqrt{B^2 + 8nC}}{4\sqrt{n(n-4)}}$.

To find the estimators of μ_x, μ_u, σ_1, μ_y, μ_v, σ_2 and ρ we replace $\sigma_{2.1}$ by $\sigma_2\sqrt{1-\rho^2}$ and θ by $\rho\frac{\sigma_2}{\sigma_1}$ in the likelihood equation (2.4.14). Equation (2.4.14) becomes:

$$\ln L \propto -2n\ln\sigma_1 - 2n\ln\sigma_2 - n\ln(1-\rho^2) - \sum_{i=1}^{n} z_{1(i)} - (b_x+1)\sum_{i=1}^{n}\ln(1+e^{-z_{1(i)}})$$
$$-\sum_{i=1}^{n} z_{2(i)} - (b_y+1)\sum_{i=1}^{n}\ln(1+e^{-z_{2(i)}}) - \sum_{i=1}^{n} w_{1(i)} - (b_u+1)\sum_{i=1}^{n}\ln(1+e^{-w_{1(i)}})$$
$$-\sum_{i=1}^{n} w_{2(i)} - (b_v+1)\sum_{i=1}^{n}\ln(1+e^{-w_{2(i)}}) . \qquad (2.4.29)$$

Here, $z_{2(i)} = \dfrac{e_{1(i)}}{\sigma_2\sqrt{1-\rho^2}} = \dfrac{y_{[i]} - \mu_y - \dfrac{\rho\sigma_2}{\sigma_1}(x_{[i]} - \mu_x)}{\sigma_2\sqrt{1-\rho^2}}$ and

$$w_{2(i)} = \frac{e_{2(i)}}{\sigma_2\sqrt{1-\rho^2}} = \frac{v_{[i]} - \mu_v - \dfrac{\rho\sigma_2}{\sigma_1}(u_{[i]} - \mu_u)}{\sigma_2\sqrt{1-\rho^2}} .$$

Differentiating equation (2.4.29) with respect to μ_x, μ_u, σ_1, μ_y, μ_v, σ_2 and ρ we get the following equations:

$$\frac{\partial\ln L}{\partial\mu_x} = \frac{n}{\sigma_1} - \frac{(b_x+1)}{\sigma_1}\sum_{i=1}^{n} g_1(z_{1(i)}) - \frac{n\rho}{\sigma_1\sqrt{1-\rho^2}} + \frac{(b_y+1)\rho}{\sigma_1\sqrt{1-\rho^2}}\sum_{i=1}^{n} g_2(z_{2(i)}) = 0 \qquad (2.4.30)$$

$$\frac{\partial\ln L}{\partial\mu_u} = \frac{n}{\sigma_1} - \frac{(b_u+1)}{\sigma_1}\sum_{i=1}^{n} h_1(w_{1(i)}) - \frac{n\rho}{\sigma_1\sqrt{1-\rho^2}} + \frac{(b_y+1)\rho}{\sigma_1\sqrt{1-\rho^2}}\sum_{i=1}^{n} h_2(w_{2(i)}) = 0$$

(2.4.31)

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{2n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_{1(i)} - \frac{(b_x+1)}{\sigma_1}\sum_{i=1}^{n} z_{1(i)} g_1(z_{1(i)}) - \frac{\rho}{\sigma_1\sqrt{1-\rho^2}}\sum_{i=1}^{n} z_{1(i)}$$

$$+\frac{(b_y+1)\rho}{\sigma_1\sqrt{1-\rho^2}}\sum_{i=1}^{n} z_{1(i)} g_2(z_{2(i)}) + \frac{1}{\sigma_1}\sum_{i=1}^{n} w_{1(i)} - \frac{(b_u+1)}{\sigma_1}\sum_{i=1}^{n} w_{1(i)} h_1(w_{1(i)})$$

$$-\frac{\rho}{\sigma_1\sqrt{1-\rho^2}}\sum_{i=1}^{n} w_{1(i)} + \frac{(b_v+1)\rho}{\sigma_1\sqrt{1-\rho^2}}\sum_{i=1}^{n} w_{1(i)} h_2(w_{2(i)}) = 0 \qquad (2.4.32)$$

$$\frac{\partial \ln L}{\partial \mu_y} = \frac{n}{\sigma_2\sqrt{1-\rho^2}} - \frac{(b_y+1)}{\sigma_2\sqrt{1-\rho^2}}\sum_{i=1}^{n} g_2(z_{2(i)}) = 0 \qquad (2.4.33)$$

$$\frac{\partial \ln L}{\partial \mu_v} = \frac{n}{\sigma_2\sqrt{1-\rho^2}} - \frac{(b_v+1)}{\sigma_2\sqrt{1-\rho^2}}\sum_{i=1}^{n} h_2(w_{2(i)}) = 0 \qquad (2.4.34)$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{2n}{\sigma_2} + \frac{1}{\sigma_2}\sum_{i=1}^{n} z_{2(i)} + \frac{\rho}{\sigma_2\sqrt{1-\rho^2}}\sum_{i=1}^{n} z_{1(i)} - \frac{(b_y+1)}{\sigma_2}\sum_{i=1}^{n} z_{2(i)} g_2(z_{2(i)})$$

$$-\frac{(b_y+1)\rho}{\sigma_2\sqrt{1-\rho^2}}\sum_{i=1}^{n} z_{1(i)} g_2(z_{2(i)}) + \frac{1}{\sigma_2}\sum_{i=1}^{n} w_{2(i)} + \frac{\rho}{\sigma_2\sqrt{1-\rho^2}}\sum_{i=1}^{n} w_{1(i)}$$

$$-\frac{(b_v+1)}{\sigma_2}\sum_{i=1}^{n} w_{2(i)} h_2(w_{2(i)}) - \frac{(b_v+1)\rho}{\sigma_2\sqrt{1-\rho^2}}\sum_{i=1}^{n} w_{1(i)} h_2(w_{2(i)}) = 0 \quad (2.4.35)$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{2n\rho}{1-\rho^2} + \frac{1}{\sqrt{1-\rho^2}}\sum_{i=1}^{n} z_{1(i)} - \frac{\rho}{1-\rho^2}\sum_{i=1}^{n} z_{2(i)} - \frac{(b_y+1)}{\sqrt{1-\rho^2}}\sum_{i=1}^{n} z_{1(i)} g_2(z_{2(i)})$$

$$+\frac{(b_y+1)\rho}{1-\rho^2}\sum_{i=1}^{n} z_{2(i)} g_2(z_{2(i)}) + \frac{1}{\sqrt{1-\rho^2}}\sum_{i=1}^{n} w_{1(i)} - \frac{\rho}{1-\rho^2}\sum_{i=1}^{n} w_{2(i)}$$

$$-\frac{(b_v+1)}{\sqrt{1-\rho^2}}\sum_{i=1}^{n} w_{1(i)} h_2(w_{2(i)}) + \frac{(b_v+1)\rho}{1-\rho^2}\sum_{i=1}^{n} w_{2(i)} h_2(w_{2(i)}) = 0 \,. (2.4.36)$$

Replacing the functions g_1, g_2 , h_1 and h_2 by their linear approximations and replacing μ_x, μ_u, σ_1 by their estimators, we solve equations (2.4.30)-(2.4.36) and get the following estimators of μ_y, μ_v, σ_2 and ρ:

8. $\hat{\mu}_y = \bar{y}_{[.]} - \hat{\theta}(\bar{x}_{[.]} - \hat{\mu}_x) - \hat{\sigma}_{2.1}\Delta_1 / m_{12}$.

9. $\hat{\mu}_v = \bar{v}_{[.]} - \hat{\theta}(\bar{u}_{[.]} - \hat{\mu}_u) - \hat{\sigma}_{2.1}\Delta_2 / m_{22}$.

10. $\hat{\sigma}_2 = \sqrt{\hat{\sigma}_{2.1}^2 + \hat{\theta}^2 \hat{\sigma}_1^2}$.

11. $\hat{\rho} = \hat{\theta}\dfrac{\hat{\sigma}_1}{\hat{\sigma}_2}$.

Computation of the MMLE:

Note that in order to compute the MML estimates we define two new variables $p_{1i} = y_i - \theta\, x_i$ and $p_{2i} = v_i - \theta\, u_i$. These two variables have the same order (in terms of order statistics) as the errors since they only differ from the errors by a constant (note that $e_{1i} = y_i - \theta\, x_i - \mu_{y/x}$ and $e_{2i} = v_i - \theta\, u_i - \mu_{v/u}$). We find the MML estimates using two iterations. In the first iteration we order our new variables using the least square estimate $\tilde{\theta}$. Meaning we order $p_{1i} = y_i - \tilde{\theta}\, x_i$ and $p_{2i} = v_i - \tilde{\theta}\, u_i$, and use this order to find the concomitants $(x_{[i]}, y_{[i]})$ and $(u_{[i]}, v_{[i]})$. We use the concomitants to compute the MML estimate of $\sigma_{2.1}$ and then of θ. In our second iteration we use the MML estimate $\hat{\theta}$ which we obtained in the first iteration to order the variables. Thus, we order $p_{1i} = y_i - \hat{\theta}\, x_i$ and $p_{2i} = v_i - \hat{\theta}\, u_i$; the order statistics will give us the new concomitants $(x_{[i]}, y_{[i]})$ and $(u_{[i]}, v_{[i]})$. We then use these concomitants to find $\hat{\sigma}_{2.1}$ and other MMLE. Thus, the MML estimates are computed in two iterations. Not more than two iterations are required for the estimates to stabilize sufficiently.

2.4.2 Properties of the MML Estimators

The MML estimators are asymptotically unbiased and efficient. They are also robust to the presence of outliers and other data anomalies (mixtures, contaminations, etc.) in the data and also to deviations from the assumed distribution. We will see these results visually later in simulation studies. The MML estimators have asymptotically the same properties as the ML estimators (Vaughan and Tiku, 2000). Thus, the asymptotic covariance matrix of the estimators $\hat{\mu}_x, \hat{\mu}_u, \hat{\sigma}_1, \hat{\mu}_{y/x}, \hat{\mu}_{v/u}, \hat{\sigma}_{2.1}$ and $\hat{\theta}$ is given by the inverse of the Fisher information matrix $I^{-1}(\mu_x, \mu_u, \sigma_1, \mu_{y/x}, \mu_{v/u}, \sigma_{2.1}, \theta)$. The elements of the Fisher information matrix I($\mu_x, \mu_u, \sigma_1, \mu_{y/x}, \mu_{v/u}, \sigma_{2.1}, \theta$) are as follows (see Appendix C for details):

1. $$I_{\mu_x\mu_x} = \frac{n}{\sigma_1^2}\left[\frac{b_x}{b_x+2}\right],$$

$$I_{\mu_x\sigma_1} = \frac{n}{\sigma_1^2}\left[\frac{b_x}{(b_x+2)}\left(\psi(b_x+1)-\psi(2)\right)\right],$$

$$I_{\mu_x\mu_u} = I_{\mu_x\mu_{y/x}} = I_{\mu_x\mu_{v/u}} = I_{\mu_x\sigma_{2.1}} = I_{\mu_x\theta} = 0.$$

2. $I_{\mu_u\mu_u} = \frac{n}{\sigma_1^2}\left[\frac{b_u}{b_u+2}\right]$,

$$I_{\mu_u\sigma_1} = \frac{n}{\sigma_1^2}\left[\frac{b_u}{b_u+2}\left(\psi(b_u+1)-\psi(2)\right)\right],$$

$$I_{\mu_u\mu_{v/u}} = I_{\mu_u\mu_{y/x}} = I_{\mu_u\sigma_{2.1}} = I_{\mu_u\theta} = 0.$$

3. $I_{\sigma_1\sigma_1} = \frac{n}{\sigma_1^2}\left[2+\frac{b_x}{b_x+2}\left(\psi'(b_x+1)+\psi'(2)+\left(\psi(b_x+1)-\psi(2)\right)^2\right)\right.$

$$\left.+\frac{b_u}{b_u+2}\left(\psi'(b_u+1)+\psi'(2)+\left(\psi(b_u+1)-\psi(2)\right)^2\right)\right],$$

$$I_{\sigma_1\mu_{y/x}} = I_{\sigma_1\mu_{v/u}} = I_{\sigma_1\sigma_{2.1}} = I_{\sigma_1\theta} = 0.$$

4. $I_{\mu_{y/x}\mu_{y/x}} = \frac{n}{\sigma_{2.1}^2}\frac{b_y}{(b_y+2)}$,

$$I_{\mu_{y/x}\mu_{v/u}} = 0,$$

$$I_{\mu_{y/x}\sigma_{2.1}} = \frac{n}{\sigma_{2.1}^2}\frac{b_y}{(b_y+2)}\left(\psi(b_y+1)-\psi(2)\right),$$

$$I_{\mu_{y/x}\theta} = \frac{n\sigma_1}{\sigma_{2.1}^2}\frac{b_y}{(b_y+2)}(\psi(b_x)-\psi(1)).$$

5. $I_{\mu_{v/u}\mu_{v/u}} = \frac{n}{\sigma_{2.1}^2}\frac{b_v}{(b_v+2)}$,

$$I_{\mu_{v/u}\sigma_{2.1}} = \frac{n}{\sigma_{2.1}^2}\frac{b_v}{(b_v+2)}\left(\psi(b_v+1)-\psi(2)\right),$$

$$I_{\mu_{v/u}\theta} = \frac{n\sigma_1}{\sigma_{2.1}^2}\frac{b_v}{(b_v+2)}\left(\psi(b_u)-\psi(1)\right).$$

6. $I_{\sigma_{2.1}\sigma_{2.1}} = \frac{n}{\sigma_{2.1}^2}\left\{2 + \frac{b_y}{b_y+2}\left(\psi'(b_y+1) + \psi'(2) + \left(\psi(b_y+1) - \psi(2)\right)^2\right)\right.$

$$\left. + \frac{b_v}{b_v+2}\left(\psi'(b_v+1) + \psi'(2) + \left(\psi(b_v+1) - \psi(2)\right)^2\right)\right\},$$

$$I_{\sigma_{2.1}\theta} = \frac{n\sigma_1}{\sigma_{2.1}^2}\left\{\frac{b_y}{b_y+2}\left[\left(\psi(b_y+1) - \psi(2)\right)\left(\psi(b_x) - \psi(1)\right)\right]\right.$$

$$\left. + \frac{b_v}{b_v+2}\left[\left(\psi(b_v+1) - \psi(2)\right)\left(\psi(b_u) - \psi(1)\right)\right]\right\}.$$

7. $I_{\theta\theta} = \frac{n\sigma_1^2}{\sigma_{2.1}^2}\left\{\frac{b_y}{b_y+2}\left[\psi'(b_x) + \psi'(1) + \left(\psi(b_x) - \psi(1)\right)^2\right]\right.$

$$\left. + \frac{b_v}{b_v+2}\left[\psi'(b_u) + \psi'(1) + \left(\psi(b_u) - \psi(1)\right)^2\right]\right\}.$$

Now, we define the Fisher information matrix I(μ_x, μ_u, σ_1, μ_y, μ_v, σ_2, ρ) for estimating μ_x, μ_u, σ_1, μ_y, μ_v, σ_2, ρ. The elements of this matrix are (see Appendix C for details):

1. $I_{\mu_x\mu_x} = \frac{n}{\sigma_1^2}\left[\frac{b_x}{b_x+2} + \frac{\rho^2}{1-\rho^2}\frac{b_y}{(b_y+2)}\right],$

$$I_{\mu_x\mu_u} = 0,$$

$$I_{\mu_x\sigma_1} = \frac{n}{\sigma_1^2}\left[\frac{b_x}{b_x+2}\left(\psi(b_x+1) - \psi(2)\right) + \frac{\rho^2}{1-\rho^2}\frac{b_y}{(b_y+2)}\left(\psi(b_x) - \psi(1)\right)\right],$$

$$I_{\mu_x\mu_y} = \frac{-n\rho}{\sigma_1\sigma_2(1-\rho^2)}\frac{b_y}{(b_y+2)},$$

$$I_{\mu_x \mu_v} = 0,$$

$$I_{\mu_x \sigma_2} = \frac{-n\rho}{\sigma_1 \sigma_2 \sqrt{1-\rho^2}} \frac{b_y}{(b_y+2)} \left[\left(\psi(b_y+1) - \psi(2) \right) + \frac{\rho}{\sqrt{1-\rho^2}} \left(\psi(b_x) - \psi(1) \right) \right],$$

$$I_{\mu_x \rho} = \frac{-n\rho}{\sigma_1 (1-\rho^2)} \frac{b_y}{(b_y+2)} \left[\left(\psi(b_x) - \psi(1) \right) - \frac{\rho}{\sqrt{1-\rho^2}} \left(\psi(b_y+1) - \psi(2) \right) \right].$$

2. $$I_{\mu_u \mu_u} = \frac{n}{\sigma_1^2} \left[\frac{b_u}{b_u+2} + \frac{\rho^2}{1-\rho^2} \frac{b_v}{(b_v+2)} \right],$$

$$I_{\mu_u \sigma_1} = \frac{n}{\sigma_1^2} \left[\frac{b_u}{b_u+2} \left(\psi(b_u+1) - \psi(2) \right) + \frac{\rho^2}{1-\rho^2} \frac{b_v}{(b_v+2)} \left(\psi(b_u) - \psi(1) \right) \right],$$

$$I_{\mu_u \mu_y} = 0,$$

$$I_{\mu_u \mu_v} = \frac{-n\rho}{\sigma_1 \sigma_2 (1-\rho^2)} \frac{b_v}{(b_v+2)},$$

$$I_{\mu_u \sigma_2} = \frac{-n\rho}{\sigma_1 \sigma_2 \sqrt{1-\rho^2}} \frac{b_v}{(b_v+2)} \left[\left(\psi(b_v+1) - \psi(2) \right) + \frac{\rho}{\sqrt{1-\rho^2}} \left(\psi(b_u) - \psi(1) \right) \right],$$

$$I_{\mu_u \rho} = \frac{-n\rho}{\sigma_1 (1-\rho^2)} \frac{b_v}{(b_v+2)} \left[\left(\psi(b_u) - \psi(1) \right) - \frac{\rho}{\sqrt{1-\rho^2}} \left(\psi(b_v+1) - \psi(2) \right) \right].$$

3. $$I_{\sigma_1 \sigma_1} = \frac{n}{\sigma_1^2} \left[2 + \frac{b_x}{b_x+2} \left(\psi'(b_x+1) + \psi'(2) + \left(\psi(b_x+1) - \psi(2) \right)^2 \right) \right.$$

$$+ \frac{\rho^2}{1-\rho^2} \frac{b_y}{(b_y+2)} \left(\psi'(b_x) + \psi'(1) + \left(\psi(b_x) - \psi(1) \right)^2 \right)$$

$$+ \frac{b_u}{b_u+2} \left(\psi'(b_u+1) + \psi'(2) + \left(\psi(b_u+1) - \psi(2) \right)^2 \right)$$

$$+\frac{\rho^2}{1-\rho^2}\frac{b_v}{(b_v+2)}\left(\psi'(b_u)+\psi'(1)+\left(\psi(b_u)-\psi(1)\right)^2\right)\Bigg],$$

$$I_{\sigma_1\mu_y}=\frac{-n\rho}{\sigma_1\sigma_2(1-\rho^2)}\frac{b_y}{(b_y+2)}\left(\psi(b_x)-\psi(1)\right),$$

$$I_{\sigma_1\mu_v}=\frac{-n\rho}{\sigma_1\sigma_2\sqrt{1-\rho^2}}\frac{b_v}{(b_v+2)}\left(\psi(b_u)-\psi(1)\right),$$

$$I_{\sigma_1\sigma_2}=\frac{-n\rho}{\sigma_1\sigma_2\sqrt{1-\rho^2}}\Bigg\{\frac{b_y}{(b_y+2)}\Big[\left(\psi(b_x)-\psi(1)\right)\left(\psi(b_y+1)-\psi(2)\right)$$

$$+\frac{\rho}{\sqrt{1-\rho^2}}\left(\psi'(b_x)+\psi'(1)+\left(\psi(b_x)-\psi(1)\right)^2\right)\Big]$$

$$+\frac{b_v}{(b_v+2)}\Big[\left(\psi(b_u)-\psi(1)\right)\left(\psi(b_v+1)-\psi(2)\right)$$

$$+\frac{\rho}{\sqrt{1-\rho^2}}\left(\psi'(b_u)+\psi'(1)+\left(\psi(b_u)-\psi(1)\right)^2\right)\Big]\Bigg\},$$

$$I_{\sigma_1\rho}=\frac{-n\rho}{\sigma_1(1-\rho^2)}\Bigg\{\frac{b_y}{(b_y+2)}\Big[\left(\psi'(b_x)+\psi'(1)+\left(\psi(b_x)-\psi(1)\right)^2\right)$$

$$-\frac{\rho}{\sqrt{1-\rho^2}}\left(\psi(b_x)-\psi(1)\right)\left(\psi(b_y+1)-\psi(2)\right)\Big]$$

$$+\frac{b_v}{(b_v+2)}\Big[\left(\psi'(b_u)+\psi'(1)+\left(\psi(b_u)-\psi(1)\right)^2\right).$$

$$-\frac{\rho}{\sqrt{1-\rho^2}}\left(\psi(b_u)-\psi(1)\right)\left(\psi(b_v+1)-\psi(2)\right)\Big]\Bigg\}.$$

4. $I_{\mu_y\mu_y}=\frac{n}{\sigma_2^2(1-\rho^2)}\frac{b_y}{(b_y+2)},$

$$I_{\mu_y \mu_v} = 0\,,$$

$$I_{\mu_y \sigma_2} = \frac{n}{\sigma_2^2\sqrt{1-\rho^2}} \frac{b_y}{(b_y+2)} \left[\left(\psi(b_y+1) - \psi(2)\right) + \frac{\rho}{\sqrt{1-\rho^2}} \left(\psi(b_x) - \psi(1)\right) \right],$$

$$I_{\mu_y \rho} = \frac{-n}{\sigma_2(1-\rho^2)} \frac{b_y}{(b_y+2)} \left[\frac{\rho^2}{\sqrt{1-\rho^2}} \left(\psi(b_y+1) - \psi(2)\right) - \left(\psi(b_x) - \psi(1)\right) \right].$$

5. $$I_{\mu_v \mu_v} = \frac{n}{\sigma_2^2(1-\rho^2)} \frac{b_v}{b_v+2}\,,$$

$$I_{\mu_v \sigma_2} = \frac{n}{\sigma_2^2\sqrt{1-\rho^2}} \frac{b_v}{(b_v+2)} \left[\left(\psi(b_v+1) - \psi(2)\right) + \frac{\rho}{\sqrt{1-\rho^2}} \left(\psi(b_u) - \psi(1)\right) \right],$$

$$I_{\mu_v \rho} = \frac{-n}{\sigma_2(1-\rho^2)} \frac{b_v}{(b_v+2)} \left[\frac{\rho^2}{\sqrt{1-\rho^2}} \left(\psi(b_v+1) - \psi(2)\right) - \left(\psi(b_u) - \psi(1)\right) \right].$$

6. $$\begin{aligned} I_{\sigma_2 \sigma_2} = \frac{n}{\sigma_2^2} \Bigg\{ 2 + \frac{b_y}{b_y+2} \Bigg[& \left(\psi'(b_y+1) + \psi'(2) + \left(\psi(b_y+1) - \psi(2)\right)^2\right) \\ & + \frac{2\rho}{\sqrt{1-\rho^2}} \left(\psi(b_x) - \psi(1)\right)\left(\psi(b_y+1) - \psi(2)\right) \\ & + \frac{\rho^2}{1-\rho^2} \left(\psi'(b_x) + \psi'(1) + \left(\psi(b_x) - \psi(1)\right)^2\right) \Bigg] \\ & + \frac{b_v}{b_v+2} \Bigg[\left(\psi'(b_v+1) + \psi'(2) + \left(\psi(b_v+1) - \psi(2)\right)^2\right) \\ & + \frac{2\rho}{\sqrt{1-\rho^2}} \left(\psi(b_u) - \psi(1)\right)\left(\psi(b_v+1) - \psi(2)\right) \\ & + \frac{\rho^2}{1-\rho^2} \left(\psi'(b_u) + \psi'(1) + \left(\psi(b_u) - \psi(1)\right)^2\right) \Bigg] \Bigg\}, \end{aligned}$$

$$I_{\sigma_2\rho} = \frac{-n}{\sigma_2(1-\rho^2)}\left\{2\rho - \frac{b_y}{b_y+2}\left[\rho\left(\psi'(b_x)+\psi'(1)+\left(\psi(b_x)-\psi(1)\right)^2\right)\right.\right.$$

$$+\frac{1-2\rho^2}{\sqrt{1-\rho^2}}\left(\psi(b_x)-\psi(1)\right)\left(\psi(b_y+1)-\psi(2)\right)$$

$$\left.-\rho\left(\psi'(b_y+1)+\psi'(2)+\left(\psi(b_y+1)-\psi(2)\right)^2\right)\right]$$

$$-\frac{b_v}{b_v+2}\left[\rho\left(\psi'(b_u)+\psi'(1)+\left(\psi(b_u)-\psi(1)\right)^2\right)\right.$$

$$+\frac{1-2\rho^2}{\sqrt{1-\rho^2}}\left(\psi(b_u)-\psi(1)\right)\left(\psi(b_v+1)-\psi(2)\right)$$

$$\left.\left.-\rho\left(\psi'(b_v+1)+\psi'(2)+\left(\psi(b_v+1)-\psi(2)\right)^2\right)\right]\right\}.$$

7. $$I_{\rho\rho} = \frac{n}{(1-\rho^2)}\left\{\frac{2\rho^2}{1-\rho^2} + \frac{b_y}{b_y+2}\left[\psi'(b_x)+\psi'(1)+\left(\psi(b_x)-\psi(1)\right)^2\right.\right.$$

$$-\frac{2\rho}{\sqrt{1-\rho^2}}\left(\psi(b_x)-\psi(1)\right)\left(\psi(b_y+1)-\psi(2)\right)$$

$$\left.+\frac{\rho^2}{1-\rho^2}\left(\psi'(b_y+1)+\psi'(2)+\left(\psi(b_y+1)-\psi(2)\right)^2\right)\right]$$

$$+\frac{b_v}{b_v+2}\left[\psi'(b_u)+\psi'(1)+\left(\psi(b_u)-\psi(1)\right)^2\right.$$

$$-\frac{2\rho}{\sqrt{1-\rho^2}}\left(\psi(b_u)-\psi(1)\right)\left(\psi(b_v+1)-\psi(2)\right)$$

$$\left.\left.+\frac{\rho^2}{1-\rho^2}\left(\psi'(b_v+1)+\psi'(2)+\left(\psi(b_v+1)-\psi(2)\right)^2\right)\right]\right\}.$$

The asymptotic covariance matrix of the estimators $\hat{\mu}_x$, $\hat{\mu}_u$, $\hat{\sigma}_1, \hat{\mu}_y$, $\hat{\mu}_v$, $\hat{\sigma}_2$, $\hat{\rho}$ is given by the inverse of the above matrix, i.e. $\Gamma^{-1}(\mu_x, \mu_u, \sigma_1, \mu_y, \mu_v, \sigma_2, \rho)$.

2.4.3 Least Square Estimators

The least square estimators are found by minimizing

$$\sum_{i=1}^{n}\left(e_{1i} - E(e_1)\right)^2,\ \sum_{i=1}^{n}\left(e_{2i} - E(e_2)\right)^2,\ \sum_{i=1}^{n}\left(x_i - E(X)\right)^2 \text{ and } \sum_{i=1}^{n}\left(u_i - E(U)\right)^2 .$$

We find the following LS estimators corrected for bias (see Appendix A for details):

1. $\tilde{\mu}_x = \bar{x} - (\psi(b_x) - \psi(1))\tilde{\sigma}_1$.

2. $\tilde{\mu}_u = \bar{u} - (\psi(b_u) - \psi(1))\tilde{\sigma}_1$.

3. $\tilde{\mu}_{y/x} = \bar{y} - \tilde{\theta}\bar{x} - (\psi(b_y) - \psi(1))\tilde{\sigma}_{2.1}$.

4. $\tilde{\mu}_{v/u} = \bar{v} - \tilde{\theta}\bar{u} - (\psi(b_v) - \psi(1))\tilde{\sigma}_{2.1}$.

5. $$\tilde{\theta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i + \sum_{i=1}^{n} (u_i - \bar{u}) v_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{n} (u_i - \bar{u})^2}.$$

6. $\tilde{\mu}_y = \bar{y} - \tilde{\theta}\tilde{\sigma}_1(\psi(b_x) - \psi(1)) - \tilde{\sigma}_{2.1}(\psi(b_y) - \psi(1))$.

7. $\tilde{\mu}_v = \bar{v} - \tilde{\theta}\tilde{\sigma}_1(\psi(b_u) - \psi(1)) - \tilde{\sigma}_{2.1}(\psi(b_v) - \psi(1))$.

8. $\tilde{\sigma}_1 = \sqrt{\frac{s_x^2 + s_u^2}{\psi'(b_x) + \psi'(b_u) + 2\psi'(1)}}$ where

$$s_x^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1) \text{ and}$$

$$s_u^2 = \sum_{i=1}^{n} (u_i - \bar{u})^2 / (n-1).$$

9. $$\tilde{\sigma}_{2.1} = \sqrt{\frac{\left(\sum_{i=1}^{n} \left(y_i - \bar{y} - \tilde{\theta}(x_i - \bar{x}) \right)^2 + \sum_{i=1}^{n} \left(v_i - \bar{v} - \tilde{\theta}(u_i - \bar{u}) \right)^2 \right) / (n-2)}{\psi'(b_y) + \psi'(b_v) + 2\psi'(1)}}.$$

10. $\tilde{\sigma}_2 = \sqrt{\tilde{\sigma}_{2.1}^2 + \tilde{\theta}^2 \tilde{\sigma}_1^2}$.

11. $\tilde{\rho} = \frac{\tilde{\theta}\tilde{\sigma}_1}{\tilde{\sigma}_2}$.

The numerical values of the psi-function ψ and its derivative ψ' are given in Tiku et al. (1999). For a comprehensive study of psi-functions and their properties, see Abramowitz and Stegun (1965).

2.4.4 Weighted Least Square Estimators

We give the weighted least square estimators for the Generalize Logistic distribution below. They are obtained by minimizing

$$\frac{1}{Var(e_1)}\sum_{i=1}^{n}\left(e_{1i}-E(e_1)\right)^2 \text{ and } \frac{1}{Var(e_2)}\sum_{i=1}^{n}\left(e_{2i}-E(e_2)\right)^2 .$$

Now, $Var(e_1)=\sigma_{2.1}^2(\psi'(b_y)+\psi'(1))$, and $Var(e_2)=\sigma_{2.1}^2(\psi'(b_v)+\psi'(1))$.

In order to make our expressions shorter, let

$c_y = \psi'(b_y)+\psi'(1)$,

$c_v = \psi'(b_v)+\psi'(1)$,

$c_x = \psi'(b_x)+\psi'(1)$ and

$c_u = \psi'(b_u)+\psi'(1)$.

We see that the above weights do not affect the estimators of the means as long as e_1 and e_2 have the same variances. So, the weighted least square estimators of μ_x, μ_u, μ_y, μ_v, $\mu_{y/x}$, $\mu_{v/u}$ remain the same as the least square estimators. However, the weighted least square estimators we obtain for σ_1,

σ_2, $\sigma_{2.1}$, θ and ρ are different. They are (see Appendix B for details): (here w denotes a weighted least square estimator)

1. $\tilde{\mu}_{xw} = \bar{x} - (\psi(b_x) - \psi(1))\tilde{\sigma}_{1w}$.

2. $\tilde{\mu}_{uw} = \bar{u} - (\psi(b_u) - \psi(1))\tilde{\sigma}_{1w}$.

3. $\tilde{\mu}_{y/xw} = \bar{y} - \tilde{\theta}_w \bar{x} - (\psi(b_y) - \psi(1))\tilde{\sigma}_{2.1w}$.

4. $\tilde{\mu}_{v/uw} = \bar{v} - \tilde{\theta}_w \bar{u} - (\psi(b_v) - \psi(1))\tilde{\sigma}_{2.1w}$.

5. $$\tilde{\theta}_w = \frac{\frac{1}{c_y}\sum_{i=1}^{n}(x_i - \bar{x})y_i + \frac{1}{c_v}\sum_{i=1}^{n}(u_i - \bar{u})v_i}{\frac{1}{c_y}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{1}{c_v}\sum_{i=1}^{n}(u_i - \bar{u})^2}.$$

6. $\tilde{\mu}_{yw} = \bar{y} - \tilde{\theta}_w \tilde{\sigma}_{1w}(\psi(b_x) - \psi(1)) - \tilde{\sigma}_{2.1w}(\psi(b_y) - \psi(1))$.

7. $\tilde{\mu}_{vw} = \bar{v} - \tilde{\theta}_w \tilde{\sigma}_{1w}(\psi(b_u) - \psi(1)) - \tilde{\sigma}_{2.1w}(\psi(b_v) - \psi(1))$.

8. $$\tilde{\sigma}_{1w} = \sqrt{\frac{\frac{1}{c_y}s_x^2 + \frac{1}{c_v}s_u^2}{\left(\frac{c_x}{c_y} + \frac{c_u}{c_v}\right)}}$$, where

$$s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1) \text{ and}$$

$$s_u^2 = \sum_{i=1}^{n} (u_i - \bar{u})^2 / (n-1) .$$

9. $$\tilde{\sigma}_{2.1w} = \sqrt{\frac{\left[\frac{1}{c_y}\sum_{i=1}^{n}\left(y_i - \bar{y} - \tilde{\theta}(x_i - \bar{x})\right)^2 + \frac{1}{c_v}\sum_{i=1}^{n}\left(v_i - \bar{v} - \tilde{\theta}(u_i - \bar{u})\right)^2\right]}{2(n-2)}} .$$

10. $$\tilde{\sigma}_{2w} = \sqrt{\tilde{\sigma}_{2.1w}^2 + \tilde{\theta}_w^2 \tilde{\sigma}_{1w}^2} .$$

11. $$\tilde{\rho}_w = \frac{\tilde{\theta}_w \tilde{\sigma}_{1w}}{\tilde{\sigma}_{2w}} .$$

In Chapter 3 we perform a simulation study and show that the MML estimators are more efficient than the least squares as well as the weighted least square estimators, even for small sample sizes. This is apparently due to non-normality of the underlying distributions.

2.4 Student's t

We now consider the situation when the marginal and conditional distributions both belong to the Student's t family. We have the same relations as before,

$$E(Y / X = x) = \mu_y + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_x) = \mu_{y/x} + \theta\, x \text{ and}$$

$$E(V / U = u) = \mu_v + \rho \frac{\sigma_2}{\sigma_1}(u - \mu_u) = \mu_{v/u} + \theta\, u .$$

We will assume the following:

(a) The marginal p.d.f of X is as follows:

$$f_x(x) = \frac{1}{\sigma_1\sqrt{r_x}\ \ B\left(1/2,\ \ r_x/2\right)}\left(1+\frac{1}{r_x}\left(\frac{x-\mu_x}{\sigma_1}\right)^2\right)^{-(r_x+1)/2}. \qquad (2.5.1)$$

(b) The p.d.f of e_1 (the p.d.f. of Y/ X = *x*) is as follows:

$$f_1(e_1) = \frac{1}{\sigma_{2.1}\sqrt{r_y}\ \ B\left(1/2,\ \ r_y/2\right)}\left(1+\frac{1}{r_y}\left(\frac{e_1}{\sigma_{2.1}}\right)^2\right)^{-(r_y+1)/2}. \qquad (2.5.2)$$

(c) The marginal p.d.f of U is as follows:

$$f_u(u) = \frac{1}{\sigma_1\sqrt{r_u}\ \ B\left(1/2,\ \ r_u/2\right)}\left(1+\frac{1}{r_u}\left(\frac{u-\mu_u}{\sigma_1}\right)^2\right)^{-(r_u+1)/2}. \qquad (2.5.3)$$

(d) The p.d.f of e_2 (the p.d.f. of V/ U = *u*) is as follows:

$$f_2(e_2) = \frac{1}{\sigma_{2.1}\sqrt{r_v}\ \ B\left(1/2,\ \ r_v/2\right)}\left(1+\frac{1}{r_v}\left(\frac{e_2}{\sigma_{2.1}}\right)^2\right)^{-(r_v+1)/2}. \qquad (2.5.4)$$

2.4.1 Modified Maximum likelihood

Given two random samples:
$(x_i, y_i), 1 \le i \le n$, from the first model and $(u_i, v_i), 1 \le i \le n$, from the second model.

As before, the likelihood function is

$\text{L} = L_x L_{y/x} L_u L_{v/u} = L_x L_{e_1} L_u L_{e_2}$.

Thus we have,

$$\text{L} \propto \prod_{i=1}^{n} \left[\frac{1}{\sigma_1}\left(1+\frac{1}{r_x}\left(\frac{x_i-\mu_x}{\sigma_1}\right)^2\right)^{-(r_x+1)/2} \frac{1}{\sigma_{2.1}}\left(1+\frac{1}{r_y}\left(\frac{e_{1i}}{\sigma_{2.1}}\right)^2\right)^{-(r_y+1)/2} \right.$$
$$\left. \frac{1}{\sigma_1}\left(1+\frac{1}{r_u}\left(\frac{u_i-\mu_u}{\sigma_1}\right)^2\right)^{-(r_u+1)/2} \frac{1}{\sigma_{2.1}}\left(1+\frac{1}{r_v}\left(\frac{e_{2i}}{\sigma_{2.1}}\right)^2\right)^{-(r_v+1)/2} \right]. \quad (2.5.5)$$

We can now express the log likelihood in terms of z_{1i}, z_{2i}, w_{1i}, w_{2i} as follows:

$$\ln L \propto -2n\ln\sigma_1 - 2n\ln\sigma_{2.1} - \frac{(r_x+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{z_{1i}^2}{r_x}\right)$$
$$-\frac{(r_y+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{z_{2i}^2}{r_y}\right) - \frac{(r_u+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{w_{1i}^2}{r_u}\right)$$
$$-\frac{(r_v+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{w_{2i}^2}{r_v}\right), \quad (2.5.6)$$

where

$$z_{1i} = \frac{x_i-\mu_x}{\sigma_1},\ z_{2i} = \frac{e_{1i}}{\sigma_{2.1}} = \sqrt{c_1(x_i)}\,\frac{y_i-\mu_{y/x}-\theta\, x_i}{\sigma_{2.1}},\ w_{1i} = \frac{u_i-\mu_u}{\sigma_1},$$

and $w_{2i} = \dfrac{e_{2i}}{\sigma_{2.1}} = \sqrt{c_2(u_i)}\,\dfrac{v_i-\mu_{v/u}-\theta\, u_i}{\sigma_{2.1}}$.

Here, c_1 and c_2 are first assumed to be equal to 1 for all values of x and u respectively.

Note that under the above assumed distributions, if we make the transformations:

$Z_1 = \dfrac{X - \mu_x}{\sigma_1}$, $\quad Z_2 = \dfrac{e_1}{\sigma_{2.1}} = \sqrt{c_1(x)}\,\dfrac{Y - \mu_{y/x} - \theta\, x}{\sigma_{2.1}}$, $\quad W_1 = \dfrac{U - \mu_u}{\sigma_1}$, and

$W_2 = \dfrac{e_2}{\sigma_{2.1}} = \sqrt{c_2(u)}\,\dfrac{V - \mu_{v/u} - \theta\, u}{\sigma_{2.1}}$ then

$$Z_1 = \frac{X - \mu_x}{\sigma_1} \sim t(r_x),$$

$$Z_2 = \sqrt{c_1(x)}\,\frac{Y - \mu_{y/x} - \theta\, x}{\sigma_{2.1}} \sim t(r_y),$$

$$W_1 = \frac{U - \mu_u}{\sigma_1} \sim t(r_u) \text{ and}$$

$$W_2 = \sqrt{c_2(u)}\,\frac{V - \mu_{v/u} - \theta\, u}{\sigma_{2.1}} \sim t(r_v).$$

Remark: In all the chapters when we refer to Student's t distribution we will be referring to the distributions given in equations (2.5.1)-(2.5.4).

In order to find estimators of the parameters μ_x, μ_u, σ_1, $\mu_{y/x}$, $\mu_{v/u}$, $\sigma_{2.1}$, and θ we differentiate equation (2.5.6) with respect to μ_x, μ_u, σ_1, $\mu_{y/x}$, $\mu_{v/u}$, $\sigma_{2.1}$, θ. The likelihood equations simplify and give the following equations:

$$\frac{\partial \ln L}{\partial \mu_x} = \frac{(r_x + 1)}{\sigma_1 r_x} \sum_{i=1}^{n} \frac{z_{1i}}{\left(1 + \frac{z_{1i}^2}{r_x}\right)} = 0 \tag{2.5.7}$$

$$\frac{\partial \ln L}{\partial \mu_u} = \frac{(r_u + 1)}{\sigma_1 r_u} \sum_{i=1}^{n} \frac{w_{1i}}{\left(1 + \frac{w_{1i}^2}{r_u}\right)} = 0 \tag{2.5.8}$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{2n}{\sigma_1} + \frac{(r_x+1)}{\sigma_1 r_x}\sum_{i=1}^{n}\frac{z_{1i}}{\left(1+\frac{z_{1i}^2}{r_x}\right)} + \frac{(r_u+1)}{\sigma_1 r_u}\sum_{i=1}^{n}\frac{w_{1i}}{\left(1+\frac{w_{1i}^2}{r_u}\right)} = 0 \tag{2.5.9}$$

$$\frac{\partial \ln L}{\partial \mu_{y/x}} = \frac{(r_y+1)}{\sigma_{2.1} r_y}\sum_{i=1}^{n}\frac{\sqrt{c_{1i}}\, z_{2i}}{\left(1+\frac{z_{2i}^2}{r_y}\right)} = 0 \tag{2.5.10}$$

$$\frac{\partial \ln L}{\partial \mu_{v/u}} = \frac{(r_v+1)}{\sigma_{2.1} r_v}\sum_{i=1}^{n}\frac{\sqrt{c_{12}}\, w_{2i}}{\left(1+\frac{w_{2i}^2}{r_v}\right)} = 0 \tag{2.5.11}$$

$$\frac{\partial \ln L}{\partial \sigma_{2.1}} = -\frac{2n}{\sigma_{2.1}} + \frac{(r_y+1)}{\sigma_{2.1} r_y}\sum_{i=1}^{n}\frac{z_{2i}^2}{\left(1+\frac{z_{2i}^2}{r_y}\right)} + \frac{(r_v+1)}{\sigma_{2.1} r_v}\sum_{i=1}^{n}\frac{w_{2i}^2}{\left(1+\frac{w_{2i}^2}{r_v}\right)} = 0 \tag{2.5.12}$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{(r_y+1)}{\sigma_{2.1} r_y}\sum_{i=1}^{n}\frac{\sqrt{c_{1i}}\;\; x_i z_{2i}}{\left(1+\frac{z_{2i}^2}{r_y}\right)} + \frac{(r_v+1)}{\sigma_{2.1} r_v}\sum_{i=1}^{n}\frac{\sqrt{c_{2i}}\;\; u_i w_{2i}}{\left(1+\frac{w_{2i}^2}{r_v}\right)} = 0. \tag{2.5.13}$$

The above likelihood equations are expressions in terms of the intractable functions:

$$g_1(z_{1i}) = \frac{z_{1i}}{\left(1+\frac{z_{1i}^2}{r_x}\right)},\; g_2(z_{2i}) = \frac{z_{2i}}{\left(1+\frac{z_{2i}^2}{r_y}\right)},$$

$$h_1(w_{1i}) = \frac{w_{1i}}{\left(1+\frac{w_{1i}^2}{r_u}\right)} \text{ and } h_2(w_{2i}) = \frac{w_{2i}}{\left(1+\frac{w_{2i}^2}{r_v}\right)},$$

and thus have no explicit solution. Again, using the method of modified maximum likelihood estimation we first order the values x_i's, e_{1i}'s, u_i's and

e_{2i} and express the log likelihood in terms of these order statistics. Here also we let $(x_{[i]}, y_{[i]})$ be the concomitants of $e_{1(i)}$, and $(u_{[i]}, v_{[i]})$ be the concomitants of $e_{2(i)}$. The log likelihood becomes:

$$\ln L \propto -2n\ln\sigma_1 - 2n\ln\sigma_{2.1} - \frac{(r_x+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{z_{1(i)}^2}{r_x}\right)$$
$$-\frac{(r_y+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{z_{2(i)}^2}{r_y}\right) - \frac{(r_u+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{w_{1(i)}^2}{r_u}\right)$$
$$-\frac{(r_v+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{w_{2(i)}^2}{r_v}\right). \qquad (2.5.14)$$

The likelihood equations become:

$$\frac{\partial \ln L}{\partial \mu_x} = \frac{(r_x+1)}{\sigma_1 r_x}\sum_{i=1}^{n} g_1(z_{1(i)}) = 0 \qquad (2.5.15)$$

$$\frac{\partial \ln L}{\partial \mu_u} = \frac{(r_u+1)}{\sigma_1 r_u}\sum_{i=1}^{n} h_1(w_{1(i)}) = 0 \qquad (2.5.16)$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{2n}{\sigma_1} + \frac{(r_x+1)}{\sigma_1 r_x}\sum_{i=1}^{n} z_{1(i)} g_1(z_{1(i)}) + \frac{(r_u+1)}{\sigma_1 r_u}\sum_{i=1}^{n} w_{1(i)} h_1(w_{1(i)}) = 0 \qquad (2.5.17)$$

$$\frac{\partial \ln L}{\partial \mu_{y/x}} = \frac{(r_y+1)}{\sigma_{2.1} r_y}\sum_{i=1}^{n} \sqrt{c_{1[i]}}\, g_2(z_{2(i)}) = 0 \qquad (2.5.18)$$

$$\frac{\partial \ln L}{\partial \mu_{v/u}} = \frac{(r_v+1)}{\sigma_{2.1} r_v}\sum_{i=1}^{n} \sqrt{c_{2[i]}}\, h_2(w_{2(i)}) = 0 \qquad (2.5.19)$$

$$\frac{\partial \ln L}{\partial \sigma_{2.1}} = -\frac{2n}{\sigma_{2.1}} + \frac{(r_y+1)}{\sigma_{2.1} r_y}\sum_{i=1}^{n} z_{2(i)} g_2(z_{2(i)}) + \frac{(r_v+1)}{\sigma_{2.1} r_v}\sum_{i=1}^{n} w_{2(i)} h_2(w_{2(i)}) = 0 \qquad (2.5.20)$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{(r_y+1)}{\sigma_{2.1} r_y}\sum_{i=1}^{n} \sqrt{c_{1[i]}}\; x_{[i]} g_2(z_{2(i)}) + \frac{(r_v+1)}{\sigma_{2.1} r_v}\sum_{i=1}^{n} \sqrt{c_{2[i]}}\; u_{[i]} h_2(w_{2(i)}) = 0. \qquad (2.5.21)$$

Now, we approximate the functions

$$g_1(z_{1(i)}) = \frac{z_{1(i)}}{\left(1+\frac{z_{1(i)}^2}{r_x}\right)} \text{ and } g_2(z_{2(i)}) = \frac{z_{2(i)}}{\left(1+\frac{z_{2(i)}^2}{r_y}\right)}$$

by the first two terms of their Taylor expansions around $t_{1(i)} = E(Z_{1(i)})$ and $t_{2(i)} = E(Z_{2(i)})$ respectively:

$$g_1(z_{1(i)}) \cong \alpha_{1i} + \beta_{1i} z_{1(i)}$$

$$g_2(z_{2(i)}) \cong \alpha_{2i} + \beta_{2i} z_{2(i)}.$$

To determine the values of the alphas and betas we write:

$$g_1(z_{1(i)}) \cong g_1(t_{1(i)}) + (z_{1(i)} - t_{1(i)}) g_1'(t_{1(i)}).$$

We obtain the following values of alphas and betas:

$$\alpha_{1i} = \frac{2t_{1(i)}^3 / r_x}{\left(1+\frac{t_{1(i)}^2}{r_x}\right)^2}, \ \beta_{1i} = \frac{1-(t_{1(i)}^2 / r_x)}{\left(1+\frac{t_{1(i)}^2}{r_x}\right)^2},$$

$$\alpha_{2i} = \frac{2t_{2(i)}^3 / r_y}{\left(1+\frac{t_{2(i)}^2}{r_y}\right)^2} \text{ and } \beta_{2i} = \frac{1-(t_{2(i)}^2 / r_y)}{\left(1+\frac{t_{2(i)}^2}{r_y}\right)^2},$$

where $t_{1(i)}$ can be obtained from the following equation:

$$\frac{1}{\sqrt{r_x}\ B(1/2,\ r_x/2)} \int_{-\infty}^{t_{1(i)}} \left(1+\frac{t^2}{r_x}\right)^{-(r_x+1)/2} dt = \frac{i}{(n+1)}, \quad 1 \le i \le n,$$

and similarly $t_{2(i)}$ can be replaced by the ith quantile of the Student's t distribution with r_y degrees of freedom.

We do the same operation for

$$h_1(w_{1(i)}) = \frac{w_{1(i)}}{\left(1+\dfrac{w_{1(i)}^2}{r_u}\right)} \text{ and } h_2(w_{2(i)}) = \frac{w_{2(i)}}{\left(1+\dfrac{w_{2(i)}^2}{r_v}\right)}$$

and approximate them by the first two terms of their Taylor series expansions around $t_{1(i)}^* = E(W_{1(i)})$ and $t_{2(i)}^* = E(W_{2(i)})$, respectively:

$$h_1(w_{1(i)}) \cong \delta_{1i} + \gamma_{1i} w_{1(i)},$$

$$h_2(w_{2(i)}) \cong \delta_{2i} + \gamma_{2i} w_{2(i)}.$$

We get the following values of deltas and gammas:

$$\delta_{1i} = \frac{2t_{1(i)}^{*3} / r_u}{\left(1+\dfrac{t_{1(i)}^{*2}}{r_u}\right)^2},\ \gamma_{1i} = \frac{1-(t_{1(i)}^{*2} / r_u)}{\left(1+\dfrac{t_{1(i)}^{*2}}{r_u}\right)^2},$$

$$\delta_{2i} = \frac{2t_{2(i)}^{*3} / r_v}{\left(1+\dfrac{t_{2(i)}^{*2}}{r_v}\right)^2} \text{ and } \gamma_{2i} = \frac{1-(t_{2(i)}^{*2} / r_v)}{\left(1+\dfrac{t_{2(i)}^{*2}}{r_v}\right)^2},$$

where $t_{1(i)}^*$ can be obtained from the following equation:

$$\frac{1}{\sqrt{r_u}\ B(1/2,\ r_u/2)} \int_{-\infty}^{t_{1(i)}^*} \left(1+\frac{t^2}{r_u}\right)^{-(r_u+1)/2} dt = \frac{i}{(n+1)},\quad 1 \le i \le n.$$

Similarly, $t_{2(i)}^*$ can be replaced by the ith quantile of the Student's t distribution with r_v degrees of freedom.

Substituting the linear approximations of the functions g_1, g_2 , h_1 and h_2 in the likelihood equations (2.5.15)-(2.5.21), we obtain the following modified maximum likelihood equations:

$$\frac{\partial \ln L^*}{\partial \mu_x} = \frac{(r_x+1)}{\sigma_1 r_x}\sum_{i=1}^{n}(\alpha_{1i}+\beta_{1i}z_{1(i)}) = 0 \tag{2.5.22}$$

$$\frac{\partial \ln L^*}{\partial \mu_u} = \frac{(r_u+1)}{\sigma_1 r_u}\sum_{i=1}^{n}(\delta_{1i}+\gamma_{1i}w_{1(i)}) = 0 \tag{2.5.23}$$

$$\frac{\partial \ln L^*}{\partial \sigma_1} = -\frac{2n}{\sigma_1}+\frac{(r_x+1)}{\sigma_1 r_x}\sum_{i=1}^{n}z_{1(i)}(\alpha_{1i}+\beta_{1i}z_{1(i)})+\frac{(r_u+1)}{\sigma_1 r_u}\sum_{i=1}^{n}w_{1(i)}(\delta_{1i}+\gamma_{1i}w_{1(i)})$$
$$= 0 \tag{2.5.24}$$

$$\frac{\partial \ln L^*}{\partial \mu_{y/x}} = \frac{(r_y+1)}{\sigma_{2.1} r_y}\sum_{i=1}^{n}\sqrt{c_{1[i]}}(\alpha_{2i}+\beta_{2i}z_{2(i)}) = 0 \tag{2.5.25}$$

$$\frac{\partial \ln L^*}{\partial \mu_{v/u}} = \frac{(r_v+1)}{\sigma_{2.1} r_v}\sum_{i=1}^{n}\sqrt{c_{2[i]}}(\delta_{2i}+\gamma_{2i}w_{2(i)}) = 0 \tag{2.5.26}$$

$$\frac{\partial \ln L^*}{\partial \sigma_{2.1}} = -\frac{2n}{\sigma_{2.1}}+\frac{(r_y+1)}{\sigma_{2.1} r_y}\sum_{i=1}^{n}z_{2(i)}(\alpha_{2i}+\beta_{2i}z_{2(i)})+\frac{(r_v+1)}{\sigma_{2.1} r_v}\sum_{i=1}^{n}w_{2(i)}(\delta_{2i}+\gamma_{2i}w_{2(i)})$$
$$= 0 \tag{2.5.27}$$

$$\frac{\partial \ln L^*}{\partial \theta} = \frac{(r_y+1)}{\sigma_{2.1} r_y}\sum_{i=1}^{n}\sqrt{c_{1[i]}}\; x_{[i]}(\alpha_{2i}+\beta_{2i}z_{2(i)})+\frac{(r_v+1)}{\sigma_{2.1} r_v}\sum_{i=1}^{n}\sqrt{c_{2[i]}}\; u_{[i]}(\delta_{2i}+\gamma_{2i}w_{2(i)})$$
$$= 0. \tag{2.5.28}$$

The above MML equations have explicit solutions. We assume that c_1 and c_2 are completely specified. In particular, we assume that $c_1(x)$ and $c_2(u)$ are both equal to 1. Solving the MML equations, we get the following MML estimators of the parameters:

1. $\hat{\mu}_x = \dfrac{\sum_{i=1}^{n} \beta_{1i} x_{(i)}}{m_{11}}$ where $m_{11} = \sum_{i=1}^{n} \beta_{1i}$.

2. $\hat{\mu}_u = \dfrac{\sum_{i=1}^{n} \gamma_{1i} u_{(i)}}{m_{21}}$ where $m_{21} = \sum_{i=1}^{n} \gamma_{1i}$.

3. $\hat{\sigma}_1^* = \dfrac{B_1 + \sqrt{B_1^{\ 2} + 8nC_1}}{4n}$ where

$$B_1 = \frac{(r_x+1)}{r_x}\sum_{i=1}^{n} \alpha_{1i} x_{(i)} + \frac{(r_u+1)}{r_u}\sum_{i=1}^{n} \delta_{1i} u_{(i)} \text{ and}$$

$$C_1 = \frac{(r_x+1)}{r_x}\sum_{i=1}^{n} \beta_{1i} (x_{(i)} - \hat{\mu}_x)^2 + \frac{(r_u+1)}{r_u}\sum_{i=1}^{n} \gamma_{1i} (u_{(i)} - \hat{\mu}_u)^2 .$$

Adjusting for the bias we get: $\hat{\sigma}_1 = \dfrac{B_1 + \sqrt{B_1^{\ 2} + 8nC_1}}{4\sqrt{n(n-2)}}$.

4. $\hat{\mu}_{y/x} = \bar{y}_{[.]} - \hat{\theta}\bar{x}_{[.]} + (\hat{\sigma}_{2.1} / m_{12})\sum_{i=1}^{n} \alpha_i$ where

$$\bar{y}_{[.]} = \frac{1}{m_{12}}\sum_{i=1}^{n} \beta_i y_{[i]} , \ \bar{x}_{[.]} = \frac{1}{m_{12}}\sum_{i=1}^{n} \beta_i x_{[i]} ,$$

$$m_{12} = \sum_{i=1}^{n} \beta_i , \ \beta_i = c_{1[i]}\beta_{2i} \text{ and } \alpha_i = \sqrt{c_{1[i]}}\alpha_{2i} .$$

5. $\hat{\mu}_{v/u} = \bar{v}_{[.]} - \hat{\theta}\bar{u}_{[.]} + (\hat{\sigma}_{2.1} / m_{22})\sum_{i=1}^{n} \delta_i$ where

$$\bar{v}_{[.]} = \frac{1}{m_{22}} \sum_{i=1}^{n} \gamma_i v_{[i]} \, , \; \bar{u}_{[.]} = \frac{1}{m_{22}} \sum_{i=1}^{n} \gamma_i u_{[i]} \, ,$$

$$m_{22} = \sum_{i=1}^{n} \gamma_i \, , \; \gamma_i = c_{2[i]} \gamma_{2i} \text{ and } \delta_i = \sqrt{c_{2[i]}} \delta_{2i} \, .$$

Note: Under the assumption that $c_1(x) = c_2(u) = 1$, $\sum_{i=1}^{n} \alpha_i = \sum_{i=1}^{n} \delta_i = 0$.

6. $\hat{\theta} = K + D \; \hat{\sigma}_{2.1}$ where

$$K = \frac{1}{S} \left[\frac{(r_y + 1)}{r_y} \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]}) y_{[i]} + \frac{(r_v + 1)}{r_v} \sum_{i=1}^{n} \gamma_i (u_{[i]} - \bar{u}_{[.]}) v_{[i]} \right],$$

$$D = \frac{1}{S} \left[\frac{(r_y + 1)}{r_y} \sum_{i=1}^{n} \alpha_i (x_{[i]} - \bar{x}_{[.]}) + \frac{(r_v + 1)}{r_v} \sum_{i=1}^{n} \delta_i (u_{[i]} - \bar{u}_{[.]}) \right],$$

$$S = \frac{(r_y + 1)}{r_y} \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2 + \frac{(r_v + 1)}{r_v} \sum_{i=1}^{n} \gamma_i (u_{[i]} - \bar{u}_{[.]})^2 \, .$$

7. $\hat{\sigma}^*_{2.1} = \dfrac{B + \sqrt{B^2 + 8nC}}{4n}$ where

$$B = \frac{(r_y + 1)}{r_y} \sum_{i=1}^{n} \alpha_i (y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]}))$$

$$+ \frac{(r_v + 1)}{r_v} \sum_{i=1}^{n} \delta_i (v_{[i]} - \bar{v}_{[.]} - K(u_{[i]} - \bar{u}_{[.]})) \, ,$$

$$C = \frac{(r_y + 1)}{r_y} \sum_{i=1}^{n} \beta_i (y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]}))^2$$

$$+ \frac{(r_v + 1)}{r_v} \sum_{i=1}^{n} \gamma_i (v_{[i]} - \bar{v}_{[.]} - K(u_{[i]} - \bar{u}_{[.]}))^2 \, .$$

Adjusting for the bias we get: $\hat{\sigma}_{2.1} = \dfrac{B + \sqrt{B^2 + 8nC}}{4\sqrt{n(n-4)}}$.

To find the estimators of μ_y, μ_v, σ_2 and ρ we replace $\sigma_{2.1}$ by $\sigma_2\sqrt{1-\rho^2}$ and θ by $\rho\dfrac{\sigma_2}{\sigma_1}$ in the likelihood equation. Equation (2.5.14) becomes:

$$\ln L \propto -2n\ln\sigma_1 - 2n\ln\sigma_2 - n\ln(1-\rho^2) - \frac{(r_x+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{z_{1(i)}^2}{r_x}\right)$$
$$-\frac{(r_y+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{z_{2(i)}^2}{r_y}\right) - \frac{(r_u+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{w_{1(i)}^2}{r_u}\right)$$
$$-\frac{(r_v+1)}{2}\sum_{i=1}^{n}\ln\left(1+\frac{w_{2(i)}^2}{r_v}\right). \tag{2.5.29}$$

Here $z_{2(i)} = \dfrac{e_{1(i)}}{\sigma_2\sqrt{1-\rho^2}} = \sqrt{c_{1i}}\,\dfrac{y_{[i]} - \mu_y - \theta\,(x_{[i]} - \mu_x)}{\sigma_2\sqrt{1-\rho^2}}$ and

$$w_{2(i)} = \frac{e_{2(i)}}{\sigma_2\sqrt{1-\rho^2}} = \sqrt{c_{2i}}\,\frac{v_{[i]} - \mu_v - \theta\,(u_{[i]} - \mu_u)}{\sigma_2\sqrt{1-\rho^2}}.$$

Differentiating equation (2.5.29) with respect to μ_x, μ_u, σ_1, μ_y, μ_v, σ_2 and ρ we get the following equations:

$$\frac{\partial \ln L}{\partial \mu_x} = \frac{(r_x+1)}{\sigma_1 r_x}\sum_{i=1}^{n} g_1(z_{1(i)}) - \frac{(r_y+1)\rho}{\sigma_1 r_y\sqrt{1-\rho^2}}\sum_{i=1}^{n}\sqrt{c_{1[i]}}\,g_2(z_{2(i)}) = 0 \tag{2.5.30}$$

$$\frac{\partial \ln L}{\partial \mu_u} = \frac{(r_u+1)}{\sigma_1 r_u}\sum_{i=1}^{n} h_1(w_{1(i)}) - \frac{(r_v+1)\rho}{\sigma_1 r_v\sqrt{1-\rho^2}}\sum_{i=1}^{n}\sqrt{c_{2[i]}}\,h_2(w_{2(i)}) = 0 \tag{2.5.31}$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{2n}{\sigma_1} + \frac{(r_x+1)}{\sigma_1 r_x}\sum_{i=1}^{n} z_{1(i)} g_1(z_{1(i)}) - \frac{(r_y+1)\rho}{\sigma_1 r_y\sqrt{1-\rho^2}}\sum_{i=1}^{n}\sqrt{c_{1[i]}}\ z_{1(i)} g_2(z_{2(i)})$$

$$+\frac{(r_u+1)}{\sigma_1 r_u}\sum_{i=1}^{n} w_{1(i)} h_1(w_{1(i)}) - \frac{(r_v+1)\rho}{\sigma_1 r_v\sqrt{1-\rho^2}}\sum_{i=1}^{n}\sqrt{c_{2[i]}}\ w_{1(i)} h_2(w_{2(i)}) = 0 \quad (2.5.32)$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{2n}{\sigma_2} + \frac{(r_y+1)}{\sigma_2^2 r_y\sqrt{1-\rho^2}}\sum_{i=1}^{n}\sqrt{c_{1[i]}}\, g_2(z_{2(i)})(y_{[i]}-\mu_y)$$

$$+\frac{(r_v+1)}{\sigma_2^2 r_v\sqrt{1-\rho^2}}\sum_{i=1}^{n}\sqrt{c_{2[i]}}\, h_2(w_{2(i)})(v_{[i]}-\mu_v) = 0 \quad (2.5.33)$$

$$\frac{\partial \ln L}{\partial \mu_y} = \frac{(r_y+1)}{\sigma_2 r_y\sqrt{1-\rho^2}}\sum_{i=1}^{n}\sqrt{c_{1[i]}}\, g_2(z_{2(i)}) = 0 \quad (2.5.34)$$

$$\frac{\partial \ln L}{\partial \mu_v} = \frac{(r_v+1)}{\sigma_2 r_v\sqrt{1-\rho^2}}\sum_{i=1}^{n}\sqrt{c_{2[i]}}\, h_2(w_{2(i)}) = 0 \quad (2.5.35)$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{2n\rho}{(1-\rho^2)} + \frac{(r_y+1)}{r_y\sqrt{1-\rho^2}}\sum_{i=1}^{n}\sqrt{c_{1[i]}}\, z_{1(i)} g_2(z_{2(i)}) - \frac{(r_y+1)\rho}{r_y(1-\rho^2)}\sum_{i=1}^{n} z_{2(i)} g_2(z_{2(i)})$$

$$+\frac{(r_v+1)}{r_v\sqrt{1-\rho^2}}\sum_{i=1}^{n}\sqrt{c_{2[i]}}\, w_{1(i)} h_2(w_{2(i)}) - \frac{(r_v+1)\rho}{r_v(1-\rho^2)}\sum_{i=1}^{n} w_{2(i)} h_2(w_{2(i)}) = 0.$$

(2.5.36)

Replacing the functions g_1, g_2 , h_1 and h_2 by their linear approximations, then replacing μ_x, μ_u, σ_1 by their estimators we can solve equations (2.5.30)-(2.5.36) and get the following estimators of μ_y, μ_v, σ_2 and ρ:

8. $\hat{\mu}_y = \bar{y}_{[.]} - \frac{\hat{\rho}\hat{\sigma}_2}{\hat{\sigma}_1}(\bar{x}_{[.]} - \hat{\mu}_x) = \bar{y}_{[.]} - \hat{\theta}(\bar{x}_{[.]} - \hat{\mu}_x) = \hat{\mu}_{y/x} + \hat{\theta}\ \hat{\mu}_x$.

9. $\hat{\mu}_v = \bar{v}_{[.]} - \dfrac{\hat{\rho}\hat{\sigma}_2}{\hat{\sigma}_1}(\bar{u}_{[.]} - \hat{\mu}_u) = \bar{v}_{[.]} - \hat{\theta}(\bar{u}_{[.]} - \hat{\mu}_u) = \hat{\mu}_{v/u} + \hat{\theta}\,\hat{\mu}_u$.

10. $\hat{\sigma}_2 = \sqrt{\hat{\sigma}_{2.1}^2 + \hat{\theta}^2\hat{\sigma}_1^2}$.

11. $\hat{\rho} = \hat{\theta}\dfrac{\hat{\sigma}_1}{\hat{\sigma}_2}$.

Remark 1: If the c's are not equal to 1, when calculating the MMLE of $\mu_{y/x}$, $\mu_{v/u}$, $\sigma_{2.1}$, and θ (and μ_y, μ_v, σ_2 and ρ), μ_x and σ_1 in $c_1(x_i)$ are replaced by their MMLE $\hat{\mu}_x$ and $\hat{\sigma}_1$. In calculating the LSE, they are replaced by $\tilde{\mu}_x$ and $\tilde{\sigma}_1$. Similarly, when calculating the MMLE, μ_u and σ_1 in $c_2(u_i)$ are replaced by $\hat{\mu}_u$ and $\hat{\sigma}_1$ and, when calculating the LSE, they are replaced by $\tilde{\mu}_u$ and $\tilde{\sigma}_1$ (See Tiku, et al., 2007).

We use the same method as before to find the concomitants $(x_{[i]}, y_{[i]})$ and $(u_{[i]}, v_{[i]})$. We use two iterations using the LSE $\tilde{\theta}$ initially.

Remark 2: Both β_{1i} and β_{2i} ($1 \le i \le n$), and similarly γ_{1i} and γ_{2i}, increase until the middle values and then decrease in a symmetric fashion. Therefore, if β_{11} and β_{21} (and similarly γ_{11} and γ_{21}) are positive, then the values of C and C_1 are positive and the values of $\hat{\sigma}_{2.1}$ and $\hat{\sigma}_1$ are both real and positive. If β_{11} is negative, which may cause C_1 to be negative, we replace α_{1i} and β_{1i} by the following values respectively:

$$\alpha_{1i}^{*} = \frac{t_{1(i)}^{3} / r_x}{\left(1 + \frac{t_{1(i)}^{2}}{r_x}\right)^2} \text{ and } \beta_{1i}^{*} = \frac{1}{\left(1 + \frac{t_{1(i)}^{2}}{r_x}\right)^2}.$$

Similarly if γ_{11} is negative we replace δ_{1i} and γ_{1i} by the values:

$$\delta_{1i}^{*} = \frac{t_{1(i)}^{*3} / r_u}{\left(1 + \frac{t_{1(i)}^{*2}}{r_u}\right)^2}, \text{ and } \gamma_{1i}^{*} = \frac{1}{\left(1 + \frac{t_{1(i)}^{*2}}{r_u}\right)^2}.$$

Also if β_{21} (or γ_{21}) is negative, which may cause C to be negative, we replace α_{2i} and β_{2i} and δ_{2i} and γ_{2i} by similar expressions:

$$\alpha_{2i}^{*} = \frac{t_{2(i)}^{3} / r_y}{\left(1 + \frac{t_{2(i)}^{2}}{r_y}\right)^2}, \ \beta_{2i}^{*} = \frac{1}{\left(1 + \frac{t_{2(i)}^{2}}{r_y}\right)^2}$$

$$\delta_{2i}^{*} = \frac{t_{2(i)}^{*3} / r_v}{\left(1 + \frac{t_{2(i)}^{*2}}{r_v}\right)^2} \text{ and } \gamma_{2i}^{*} = \frac{1}{\left(1 + \frac{t_{2(i)}^{*2}}{r_v}\right)^2}.$$

These operations do not alter the asymptotic properties of the MML estimators for reasons stated before.

2.5.2 Properties of the MML Estimators

The MML estimators we derive here have the same properties as those mentioned in section 2.4.2. Thus, the asymptotic covariance matrix of the estimators $\hat{\mu}_x$, $\hat{\mu}_u$, $\hat{\sigma}_1$, $\hat{\mu}_{y/x}$, $\hat{\mu}_{v/u}$, $\hat{\sigma}_{2.1}, \hat{\theta}$ is given by the inverse of the

Fisher information matrix $I^{-1}(\mu_x, \mu_u, \sigma_1, \mu_{y/x}, \mu_{v/u}, \sigma_{2.1}, \theta)$. The elements of the Fisher information matrix $I(\mu_x, \mu_u, \sigma_1, \mu_{y/x}, \mu_{v/u}, \sigma_{2.1}, \theta)$ are as follows (see Appendix C for details):

Information Matrix for μ_x, μ_u, $\sigma_1, \mu_{y/x}$, $\mu_{v/u}$, $\sigma_{2.1}$, θ

1. $I_{\mu_x\mu_x} = \frac{n}{\sigma_1^2}\left[\frac{r_x+1}{r_x+3}\right]$,

$I_{\mu_x\mu_u} = I_{\mu_x\sigma_1} = I_{\mu_x\mu_{y/x}} = I_{\mu_x\mu_{v/u}} = I_{\mu_x\sigma_{2.1}} = I_{\mu_x\theta} = 0$.

2. $I_{\mu_u\mu_u} = \frac{n}{\sigma_1^2}\left[\frac{r_u+1}{r_u+3}\right]$,

$I_{\mu_u\sigma_1} = I_{\mu_u\mu_{v/u}} = I_{\mu_u\mu_{y/x}} = I_{\mu_u\sigma_{2.1}} = I_{\mu_u\theta} = 0$.

3. $I_{\sigma_1\sigma_1} = \frac{n}{\sigma_1^2}\left[2 + \frac{r_x-3}{r_x+3} + \frac{r_u-3}{r_u+3}\right]$,

$I_{\sigma_1\mu_{y/x}} = I_{\sigma_1\mu_{v/u}} = I_{\sigma_1\sigma_{2.1}} = I_{\sigma_1\theta} = 0$.

4. $I_{\mu_{y/x}\mu_{y/x}} = \frac{n}{\sigma_{2.1}^2}\frac{(r_y+1)}{(r_y+3)}$,

$I_{\mu_{y/x}\mu_{v/u}} = I_{\mu_{y/x}\sigma_{2.1}} = 0$,

$I_{\mu_{y/x}\theta} = \frac{n\mu_x}{\sigma_{2.1}^2}\frac{(r_y+1)}{(r_y+3)}$.

5. $I_{\mu_{v/u}\mu_{v/u}} = \frac{n}{\sigma_{2.1}^2}\frac{(r_v+1)}{(r_v+3)}$,

$I_{\mu_{v/u}\sigma_{2.1}} = 0,$

$$I_{\mu_{v/u}\theta} = \frac{n\mu_u}{\sigma_{2.1}^2}\frac{(r_v+1)}{(r_v+3)}.$$

6. $I_{\sigma_{2.1}\sigma_{2.1}} = \frac{n}{\sigma_{2.1}^2}\left[2+\frac{r_y-3}{r_y+3}+\frac{r_v-3}{r_v+3}\right]$, $I_{\sigma_{2.1}\theta}$ =0.

7. $I_{\theta\theta} = \frac{n}{\sigma_{2.1}^2}\left\{\frac{(r_y+1)}{(r_y+3)}\frac{\left[r_x(\sigma_1^2+\mu_x^2)-2\mu_x^2\right]}{(r_x-2)}+\frac{(r_v+1)}{(r_v+3)}\frac{\left[r_u(\sigma_1^2+\mu_u^2)-2\mu_u^2\right]}{(r_u-2)}\right\}.$

Information Matrix for μ_x , μ_u , σ_1, μ_y , μ_v , σ_2 , ρ

1. $I_{\mu_x\mu_x} = \frac{n}{\sigma_1^2}\left[\frac{r_x+1}{r_x+3}+\frac{\rho^2}{(1-\rho^2)}\frac{(r_y+1)}{(r_y+3)}\right],$

$$I_{\mu_x\mu_y} = \frac{-n\rho}{\sigma_1\sigma_2(1-\rho^2)}\frac{(r_y+1)}{(r_y+3)},$$

$$I_{\mu_x\mu_u} = I_{\mu_x\sigma_1} = I_{\mu_x\mu_v} = I_{\mu_x\sigma_2} = I_{\mu_x\rho} = 0.$$

2. $I_{\mu_u\mu_u} = \frac{n}{\sigma_1^2}\left[\frac{r_u+1}{r_u+3}+\frac{\rho^2}{(1-\rho^2)}\frac{(r_v+1)}{(r_v+3)}\right],$

$$I_{\mu_u\mu_v} = \frac{-n\rho}{\sigma_1\sigma_2(1-\rho^2)}\frac{(r_v+1)}{(r_v+3)},$$

$$I_{\mu_u\sigma_1} = I_{\mu_u\mu_y} = I_{\mu_u\sigma_2} = I_{\mu_u\rho} = 0.$$

3. $I_{\sigma_1\sigma_1} = \frac{n}{\sigma_1^2}\left\{2+\frac{r_x-3}{r_x+3}+\frac{r_u-3}{r_u+3}+\frac{\rho^2}{(1-\rho^2)}\left(\frac{r_x}{(r_x-2)}\frac{(r_y+1)}{(r_y+3)}+\frac{r_u}{(r_u-2)}\frac{(r_v+1)}{(r_v+3)}\right)\right\},$

$$I_{\sigma_1\mu_y} = I_{\sigma_1\mu_v} = 0,$$

$$I_{\sigma_1\sigma_2} = \frac{-n\rho^2}{\sigma_1\sigma_2(1-\rho^2)}\left(\frac{r_x}{(r_x-2)}\frac{(r_y+1)}{(r_y+3)} + \frac{r_u}{(r_u-2)}\frac{(r_v+1)}{(r_v+3)}\right),$$

$$I_{\sigma_1\rho} = \frac{-n\rho}{\sigma_1(1-\rho^2)}\left(\frac{r_x}{(r_x-2)}\frac{(r_y+1)}{(r_y+3)} + \frac{r_u}{(r_u-2)}\frac{(r_v+1)}{(r_v+3)}\right).$$

4. $I_{\mu_y\mu_y} = \dfrac{n}{\sigma_2^2(1-\rho^2)}\dfrac{(r_y+1)}{(r_y+3)}$, $I_{\mu_y\mu_v} = I_{\mu_y\sigma_2} = I_{\mu_y\rho} = 0.$

5. $I_{\mu_v\mu_v} = \dfrac{n}{\sigma_2^2(1-\rho^2)}\dfrac{(r_v+1)}{(r_v+3)}$, $I_{\mu_v\sigma_2} = I_{\mu_v\rho} = 0.$

6. $I_{\sigma_2\sigma_2} = \dfrac{n}{\sigma_2^2}\left\{2 + \dfrac{r_y-3}{r_y+3} + \dfrac{r_v-3}{r_v+3} + \dfrac{\rho^2}{(1-\rho^2)}\left(\dfrac{r_x}{(r_x-2)}\dfrac{(r_y+1)}{(r_y+3)} + \dfrac{r_u}{(r_u-2)}\dfrac{(r_v+1)}{(r_v+3)}\right)\right\},$

$$I_{\sigma_2\rho} = \frac{-n\rho}{\sigma_2(1-\rho^2)}\left\{2 + \frac{r_y-3}{r_y+3} + \frac{r_v-3}{r_v+3} - \frac{r_x}{(r_x-2)}\frac{(r_y+1)}{(r_y+3)} - \frac{r_u}{(r_u-2)}\frac{(r_v+1)}{(r_v+3)}\right\}.$$

7. $I_{\rho\rho} = \dfrac{n}{(1-\rho^2)}\left\{\dfrac{\rho^2}{(1-\rho^2)}\left(2 + \dfrac{r_y-3}{r_y+3} + \dfrac{r_v-3}{r_v+3}\right) + \dfrac{r_x}{(r_x-2)}\dfrac{(r_y+1)}{(r_y+3)} + \dfrac{r_u}{(r_u-2)}\dfrac{(r_v+1)}{(r_v+3)}\right\}.$

2.5.3 Least Square Estimators

We minimize:

1. $\sum_{i=1}^{n} e_{1i}^2$ and $\sum_{i=1}^{n} e_{2i}^2$

2. $\sum_{i=1}^{n}(x_i - \mu_x)^2$ and $\sum_{i=1}^{n}(u_i - \mu_u)^2$,

$e_{1i} = \sqrt{c_{1i}}\,(y_i - \mu_{y/x} - \theta\, x_i)$,

$e_{2i} = \sqrt{c_{2i}}\,(v_i - \mu_{v/u} - \theta\, u_i)$.

We find the following LSE corrected for bias (see Appendix A for details):

1. $\tilde{\mu}_x = \bar{x}$.

2. $\tilde{\mu}_u = \bar{u}$.

3. $\tilde{\mu}_{y/x} = \bar{y}_{(.)} - \tilde{\theta}\bar{x}_{(.)}$, where

$$\bar{y}_{(.)} = \frac{\sum_{i=1}^{n} c_{1i}\, y_i}{\sum_{i=1}^{n} c_{1i}} \text{ and } \bar{x}_{(.)} = \frac{\sum_{i=1}^{n} c_{1i}\, x_i}{\sum_{i=1}^{n} c_{1i}}.$$

4. $\tilde{\mu}_{v/u} = \bar{v}_{(.)} - \tilde{\theta}\bar{u}_{(.)}$, where

$$\bar{v}_{(.)} = \frac{\sum_{i=1}^{n} c_{2i}\, v_i}{\sum_{i=1}^{n} c_{2i}} \text{ and } \bar{u}_{(.)} = \frac{\sum_{i=1}^{n} c_{2i}\, u_i}{\sum_{i=1}^{n} c_{2i}}.$$

5. $$\tilde{\theta} = \frac{\sum_{i=1}^{n} c_{1i}(x_i - \bar{x}_{(.)}) y_i + \sum_{i=1}^{n} c_{2i}(u_i - \bar{u}_{(.)}) v_i}{\sum_{i=1}^{n} c_{1i}(x_i - \bar{x}_{(.)})^2 + \sum_{i=1}^{n} c_{2i}(u_i - \bar{u}_{(.)})^2}.$$

6. $$\tilde{\mu}_y = \bar{y}_{(.)} - \frac{\tilde{\rho}\tilde{\sigma}_2}{\tilde{\sigma}_1}(\bar{x}_{(.)} - \tilde{\mu}_x) = \bar{y}_{(.)} - \tilde{\theta}(\bar{x}_{(.)} - \tilde{\mu}_x) = \tilde{\mu}_{y/x} + \tilde{\theta}\bar{x}.$$

7. $$\tilde{\mu}_v = \bar{v}_{(.)} - \frac{\tilde{\rho}\tilde{\sigma}_2}{\tilde{\sigma}_1}(\bar{u}_{(.)} - \tilde{\mu}_u) = \bar{v}_{(.)} - \tilde{\theta}(\bar{u}_{(.)} - \tilde{\mu}_u) = \tilde{\mu}_{v/u} + \tilde{\theta}\bar{u}.$$

8. $$\tilde{\sigma}_1 = \sqrt{\frac{s_x^2 + s_u^2}{\frac{r_x}{r_x - 2} + \frac{r_u}{r_u - 2}}},$$ where

$$s_x^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1),$$ and

$$s_u^2 = \sum_{i=1}^{n} (u_i - \bar{u})^2 / (n-1).$$

9. $$\tilde{\sigma}_{2.1} = \sqrt{\frac{\left[\sum_{i=1}^{n} c_{1i}\left(y_i - \bar{y}_{(.)} - \tilde{\theta}(x_i - \bar{x}_{(.)})\right)^2 + \sum_{i=1}^{n} c_{2i}\left(v_i - \bar{v}_{(.)} - \tilde{\theta}(u_i - \bar{u}_{(.)})\right)^2\right]}{(n-2)\left[\frac{r_y}{r_y - 2} + \frac{r_v}{r_v - 2}\right]}}.$$

10. $$\tilde{\sigma}_2 = \sqrt{\tilde{\sigma}_{2.1}^2 + \tilde{\theta}^2 \tilde{\sigma}_1^2}.$$

11. $\tilde{\rho} = \dfrac{\tilde{\theta}\tilde{\sigma}_1}{\tilde{\sigma}_2}$.

2.5.4 Weighted Least Square Estimators

We minimize:

$$\frac{1}{Var(e_{1i})}\sum_{i=1}^{n} e_{1i}^{\;2} \text{ and } \frac{1}{Var(e_{2i})}\sum_{i=1}^{n} e_{2i}^{\;2} .$$

Let

$$c_y = \frac{r_y}{r_y - 2},\ c_x = \frac{r_x}{r_x - 2},\ c_u = \frac{r_u}{r_u - 2}, \text{ and } c_v = \frac{r_v}{r_v - 2}.$$

We find the following weighted LSE corrected for bias (See Appendix B for details):

1. $\tilde{\mu}_{xw} = \bar{x}$.

2. $\tilde{\mu}_{uw} = \bar{u}$.

3. $\tilde{\mu}_{y/xw} = \bar{y}_{(.)} - \tilde{\theta}_w \bar{x}_{(.)}$.

4. $\tilde{\mu}_{v/uw} = \bar{v}_{(.)} - \tilde{\theta}_w \bar{u}_{(.)}$.

5. $$\tilde{\theta}_w = \frac{\frac{1}{c_y}\sum_{i=1}^{n} c_{1i}(x_i - \bar{x}_{(.)})y_i + \frac{1}{c_v}\sum_{i=1}^{n} c_{2i}(u_i - \bar{u}_{(.)})v_i}{\frac{1}{c_y}\sum_{i=1}^{n} c_{1i}(x_i - \bar{x}_{(.)})^2 + \frac{1}{c_v}\sum_{i=1}^{n} c_{2i}(u_i - \bar{u}_{(.)})^2}.$$

6. $$\tilde{\mu}_{yw} = \bar{y}_{(.)} - \frac{\tilde{\rho}_w \tilde{\sigma}_{2w}}{\tilde{\sigma}_{1w}}(\bar{x}_{(.)} - \tilde{\mu}_{xw}) = \bar{y}_{(.)} - \tilde{\theta}_w(\bar{x}_{(.)} - \tilde{\mu}_{xw}) = \tilde{\mu}_{y/xw} + \tilde{\theta}_w \bar{x}.$$

7. $$\tilde{\mu}_{vw} = \bar{v}_{(.)} - \frac{\tilde{\rho}_w \tilde{\sigma}_{2w}}{\tilde{\sigma}_{1w}}(\bar{u}_{(.)} - \tilde{\mu}_{uw}) = \bar{v}_{(.)} - \tilde{\theta}_w(\bar{u}_{(.)} - \tilde{\mu}_{uw}) = \tilde{\mu}_{v/uw} + \tilde{\theta}_w \bar{u}.$$

8. $$\tilde{\sigma}_{1w} = \sqrt{\frac{\frac{1}{c_y}s_x^2 + \frac{1}{c_v}s_u^2}{\left(\frac{c_x}{c_y} + \frac{c_u}{c_v}\right)}}$$, where

$$s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1) \text{ and}$$

$$s_u^2 = \sum_{i=1}^{n}(u_i - \bar{u})^2/(n-1).$$

9. $$\tilde{\sigma}_{2.1w} = \sqrt{\frac{\left[\frac{1}{c_y}\sum_{i=1}^{n} c_{1i}\left(y_i - \bar{y}_{(.)} - \tilde{\theta}(x_i - \bar{x}_{(.)})\right)^2 + \frac{1}{c_v}\sum_{i=1}^{n} c_{2i}\left(v_i - \bar{v}_{(.)} - \tilde{\theta}(u_i - \bar{u}_{(.)})\right)^2\right]}{2(n-2)}}.$$

10. $$\tilde{\sigma}_{2w} = \sqrt{\tilde{\sigma}_{2.1w}^2 + \tilde{\theta}_w^2 \tilde{\sigma}_{1w}^2}.$$

11. $\tilde{\rho}_w = \dfrac{\tilde{\theta}_w \tilde{\sigma}_{1w}}{\tilde{\sigma}_{2w}}$.

We will see in Chapter 3 that the MML estimators in this case also are much more efficient than the least squares and the weighted least square estimators.

Note: The estimators given in this chapter were derived assuming equal sample sizes, i.e. $n_1 = n_2 = n$. The MML and LS estimators in the case when $n_1 \neq n_2$ are given in Appendix H.

CHAPTER 3

SIMULATION RESULTS

In a simulation study we show now that the MML estimators developed in Chapter 2 are more efficient than the LS estimators and the weighted LS estimators. We also show that the MML estimators are more robust to deviations from the assumed distribution and also to the presence of outliers and other anomalies in the data. We perform our study assuming the Generalized Logistic distribution and the Student's t-distribution. Similar results for the efficiency and robustness of the MML estimators can be shown to be true for other distributions as well. In our study, we consider various values of the sample size n (20, 30, 50, and 100). The simulations we perform are based on [100000/n] Monte Carlo runs. Without any loss of generality, we carry out our study assuming that the actual values of the parameters are:

$$\mu_x = 0, \mu_u = 0,\ \mu_y = 0,\ \mu_v = 0, \sigma_1 = 1, \text{ and } \sigma_2 = 1.$$

3.1 Generalized Logistic

Here we consider different values of b_x, b_y, b_u, and b_v and $\rho = 0.5$, 0.2 and $\rho = 0.9$. The values for the negative values of ρ are exactly similar to the positive values and are not, therefore, reported.

3.1.1 Efficiency

Efficiency is a very important attribute of an estimator. We would like our estimator to be fully efficient which means that its variance is, at any rate, almost equal to the MVB. In the tables below we give values of the variances of our MML estimators (multiplied by n) and also we give n times the diagonal elements of the inverse of the Fisher information matrix. We show that even for small n, the variances of our estimators are close to the corresponding diagonal elements. This implies that the MMLE are highly efficient.

Table 3.1.1 Simulated variances of the MMLE vs. the diagonal elements of I^{-1} (GL), $b_x = b_y = b_u = b_v = 1$.

n=20	μ_x	μ_y	μ_u	μ_v	$\mu_{y/x}$	$\mu_{v/u}$	σ_1	σ_2	$\sigma_{2.1}$	ρ
n*V_MML	3.06	3.05	3.09	3.09	2.36	2.34	0.41	0.38	0.34	0.32
n*Fisher	3.00	3.00	3.00	3.00	2.25	2.25	0.35	0.30	0.26	0.29
n=30										
n*V_MML	3.06	3.10	3.08	2.92	2.34	2.20	0.40	0.35	0.30	0.32
n*Fisher	3.00	3.00	3.00	3.00	2.25	2.25	0.35	0.30	0.26	0.29
n=50										
n*V_MML	3.27	3.04	2.90	3.08	2.16	2.39	0.36	0.33	0.31	0.30
n*Fisher	3.00	3.00	3.00	3.00	2.25	2.25	0.35	0.30	0.26	0.29
n=100										
n*V_MML	3.04	2.87	3.26	2.96	2.28	2.23	0.36	0.32	0.28	0.32
n*Fisher	3.00	3.00	3.00	3.00	2.25	2.25	0.35	0.30	0.26	0.29

The seemingly unexpected result that some simulated variances are a little smaller than the corresponding diagonal elements can be due to simulation errors. Theoretically, however, this is possible in bivariate and multivariate situations if the estimators are correlated.

Table 3.1.2 Simulated variances of the MMLE vs. the diagonal elements of I^{-1} (GL), $b_x = b_y = b_u = b_v = 0.5$.

n=20	μ_x	μ_y	μ_u	μ_v	$\mu_{y/x}$	$\mu_{v/u}$	σ_1	σ_2	$\sigma_{2.1}$	ρ
n*V_MML	5.16	5.88	5.32	5.87	4.70	4.61	0.44	0.39	0.36	0.30
n*Fisher	5.07	5.72	5.07	5.72	4.08	4.08	0.38	0.31	0.28	0.28
n=30										
n*V_MML	5.21	5.93	5.16	5.41	4.63	4.24	0.43	0.36	0.33	0.29
n*Fisher	5.07	5.72	5.07	5.72	4.08	4.08	0.38	0.31	0.28	0.28
n=50										
n*V_MML	5.45	5.66	4.80	5.66	4.27	4.58	0.40	0.34	0.33	0.27
n*Fisher	5.07	5.72	5.07	5.72	4.08	4.08	0.38	0.31	0.28	0.28
n=100										
n*V_MML	5.10	5.13	5.49	5.47	4.17	4.21	0.39	0.33	0.30	0.30
n*Fisher	5.07	5.72	5.07	5.72	4.08	4.08	0.38	0.31	0.28	0.28

An estimator must also be unbiased. Note that the MML estimators are self-bias correcting, whereas the LS estimators will need bias correction if the mean of the random error is not zero. If the variance of the error term is not $\sigma_{2.1}^2$, its estimator $\tilde{\sigma}_{2.1}^2$ needs scale adjustments, and the same applies to $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$. We compare in the tables that follow the efficiencies of the MML, the LS and the weighted least square estimators.

The simulated values are given in the tables below. The values of the expected values and variances of the MML, LS and weighted least square

(WLS) estimators are given. We also give the relative efficiencies of the LS, namely, $100\times\frac{Var(\hat{\tau})}{Var(\tilde{\tau})}$, and also the relative efficiencies of the WLS estimators as compared to the MML estimators.

Note that the Generalized Logistic distribution is negatively skewed if the shape parameter b<1, symmetric if b = 1, and positively skewed if b>1. We start by choosing the following values of b's: $b_x = 0.5, b_y = 0.5, b_u = 0.5$, and $b_v = 0.5$. Thus, in the first case we consider the situation where all the distributions are negatively skewed. The results are given in the tables that follow.

We will use the following abbreviations throughout this chapter:
E_MMLE : Expected values of the MMLE
E_LSE: Expected values of the LSE
E_WLSE: Expected values of the WLSE
V_MMLE: Variances of the MMLE
V_LSE: Variances of the LSE
V_WLSE: Variances of the WLSE
RE(MML/LS): Relative efficiency of LS to MML.
RE(MML/WLS): Relative efficiency of WLS to MML.

Table 3.1.3 Simulation results (GL), $b_x = b_y = b_u = b_v = 0.5$, ρ =0.5.

b_x=0.5, b_y=0.5, b_u=0.5, b_v=0.5, ρ=0.5.							
n = 20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	-0.005	-0.008	0.005	0.001	1.047	1.099	0.477
E_LS	-0.019	-0.017	-0.011	-0.016	0.983	1.011	0.485
E_WLS	-0.019	-0.017	-0.011	-0.016	0.983	1.011	0.485
V_MML	0.258	0.294	0.266	0.293	0.022	0.019	0.015
V_LS	0.304	0.35	0.31	0.347	0.026	0.022	0.019
V_WLS	0.304	0.35	0.31	0.347	0.026	0.022	0.019
RE (MML/LS)	85	84	86	85	85	88	79
RE (MML/WLS)	85	84	86	85	85	88	79
n = 30							
E_MML	-0.003	0.004	-0.003	-0.01	1.03	1.062	0.485
E_LS	-0.008	0.001	-0.019	-0.022	0.989	1.007	0.491
E_WLS	-0.008	0.001	-0.019	-0.022	0.989	1.007	0.491
V_MML	0.174	0.198	0.172	0.181	0.014	0.012	0.01
V_LS	0.205	0.234	0.203	0.218	0.018	0.015	0.013
V_WLS	0.205	0.234	0.203	0.218	0.018	0.015	0.013
RE (MML/LS)	85	85	84	83	79	80	77
RE (MML/WLS)	85	85	84	83	79	80	77
n=100							
E_MML	-0.004	0.003	-0.009	-0.014	1.006	1.016	0.494
E_LS	-0.01	-0.001	-0.009	-0.015	0.995	1.001	0.495
E_WLS	-0.01	-0.001	-0.009	-0.015	0.995	1.001	0.495
V_MML	0.051	0.051	0.055	0.055	0.004	0.003	0.003
V_LS	0.062	0.061	0.067	0.069	0.006	0.004	0.004
V_WLS	0.062	0.061	0.067	0.069	0.006	0.004	0.004
RE (MML/LS)	82	84	82	79	71	74	77
RE (MML/WLS)	82	84	82	79	71	74	77

Table 3.1.4 Simulation results (GL), b_x=0.5, b_y=3, b_u=0.5, b_v=3, ρ=0.5.

b_x=0.5, b_y=3, b_u=0.5, b_v=3, ρ=0.5.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	-0.005	-0.042	0.005	-0.038	1.047	1.082	0.484
E_LS	-0.019	-0.017	-0.011	-0.013	0.983	1.003	0.491
E_WLS	-0.019	-0.017	-0.011	-0.013	0.983	1.003	0.491
V_MML	0.258	0.153	0.266	0.154	0.022	0.015	0.008
V_LS	0.304	0.174	0.31	0.173	0.026	0.016	0.01
V_WLS	0.304	0.174	0.31	0.173	0.026	0.016	0.01
RE (MML/LS)	85	88	86	89	85	92	82
RE (MML/WLS)	85	88	86	89	85	92	82
n=30							
E_MML	-0.003	-0.02	-0.003	-0.029	1.03	1.052	0.49
E_LS	-0.008	-0.002	-0.019	-0.017	0.989	1.002	0.495
E_WLS	-0.008	-0.002	-0.019	-0.017	0.989	1.002	0.495
V_MML	0.174	0.1	0.172	0.095	0.014	0.009	0.005
V_LS	0.205	0.116	0.203	0.111	0.018	0.01	0.007
V_WLS	0.205	0.116	0.203	0.111	0.018	0.01	0.007
RE (MML/LS)	85	86	84	86	79	84	79
RE (MML/WLS)	85	86	84	86	79	84	79
n=100							
E_MML	-0.004	-0.006	-0.009	-0.016	1.006	1.014	0.496
E_LS	-0.01	-0.001	-0.009	-0.009	0.995	0.999	0.498
E_WLS	-0.01	-0.001	-0.009	-0.009	0.995	0.999	0.498
V_MML	0.051	0.028	0.055	0.029	0.004	0.002	0.002
V_LS	0.062	0.033	0.067	0.035	0.006	0.003	0.002
V_WLS	0.062	0.033	0.067	0.035	0.006	0.003	0.002
RE (MML/LS)	82	84	82	83	71	77	77
RE (MML/WLS)	82	84	82	83	71	77	77

Table 3.1.5 Simulation results (GL), b_x =6, b_y =0.5, b_u =6, b_v =0.5, ρ =0.5.

b_x =6, b_y =0.5, b_u =6, b_v =0.5, ρ =0.5.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	-0.002	0.003	0.001	0.007	1.031	1.108	0.447
E_LS	0.036	0.01	0.041	0.009	0.984	1.031	0.461
E_WLS	0.036	0.01	0.041	0.009	0.984	1.031	0.461
V_MML	0.108	0.531	0.102	0.552	0.018	0.028	0.04
V_LS	0.145	0.639	0.14	0.659	0.024	0.034	0.051
V_WLS	0.145	0.639	0.14	0.659	0.024	0.034	0.051
RE (MML/LS)	74	83	73	84	75	84	78
RE (MML/WLS)	74	83	73	84	75	84	78
n=30							
E_MML	0.002	0.017	-0.002	0.002	1.02	1.066	0.463
E_LS	0.027	0.014	0.025	-0.005	0.988	1.019	0.473
E_WLS	0.027	0.014	0.025	-0.005	0.988	1.019	0.473
V_MML	0.068	0.356	0.071	0.353	0.012	0.018	0.027
V_LS	0.096	0.435	0.097	0.441	0.016	0.023	0.035
V_WLS	0.096	0.435	0.097	0.441	0.016	0.023	0.035
RE (MML/LS)	72	82	73	80	74	78	77
RE (MML/WLS)	72	82	73	80	74	78	77
n=100							
E_MML	0.002	0.019	0.003	0.005	1.004	1.016	0.486
E_LS	0.011	0.016	0.01	0	0.994	1.004	0.489
E_WLS	0.011	0.016	0.01	0	0.994	1.004	0.489
V_MML	0.02	0.105	0.021	0.104	0.003	0.005	0.008
V_LS	0.03	0.129	0.029	0.131	0.005	0.007	0.01
V_WLS	0.03	0.129	0.029	0.131	0.005	0.007	0.01
RE (MML/LS)	67	81	72	79	68	76	79
RE (MML/WLS)	67	81	72	79	68	76	79

Table 3.1.6 Simulation results (GL), b_x=0.5, b_y=3, b_u=6, b_v=0.5, ρ=0.5.

b_x=0.5, b_y=3, b_u=6, b_v=0.5, ρ=0.5.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	-0.008	-0.042	-0.011	-0.003	1.039	1.086	0.477
E_LS	-0.025	-0.025	0.054	0.018	0.979	1.01	0.485
E_WLS	-0.032	-0.025	0.065	0.021	0.974	1.006	0.483
V_MML	0.258	0.164	0.107	0.283	0.02	0.016	0.011
V_LS	0.279	0.219	0.226	0.352	0.033	0.025	0.017
V_WLS	0.277	0.182	0.292	0.366	0.042	0.02	0.015
RE (MML/LS)	92	75	47	80	61	66	64
RE (MML/WLS)	93	90	37	77	48	84	73
n=30							
E_MML	-0.005	-0.02	-0.01	-0.01	1.025	1.054	0.485
E_LS	-0.016	-0.009	0.037	0.005	0.983	1.005	0.489
E_WLS	-0.021	-0.01	0.045	0.007	0.98	1.004	0.488
V_MML	0.174	0.106	0.073	0.182	0.012	0.01	0.007
V_LS	0.188	0.145	0.159	0.227	0.022	0.017	0.011
V_WLS	0.188	0.121	0.204	0.239	0.029	0.013	0.01
RE (MML/LS)	93	73	46	80	55	60	61
RE (MML/WLS)	92	88	36	76	44	76	69
n=100							
E_MML	-0.005	-0.006	0.002	-0.003	1.004	1.015	0.494
E_LS	-0.01	-0.005	0.007	-0.002	0.995	1.001	0.496
E_WLS	-0.01	-0.003	0.008	-0.004	0.995	1.001	0.496
V_MML	0.051	0.03	0.021	0.054	0.004	0.003	0.002
V_LS	0.055	0.042	0.049	0.068	0.007	0.005	0.004
V_WLS	0.056	0.035	0.064	0.072	0.009	0.004	0.003
RE (MML/LS)	92	72	43	79	50	56	62
RE (MML/WLS)	91	85	33	75	38	72	65

Table 3.1.7 Simulation results (GL), $b_x = b_y = b_u = b_v = 1$, $\rho = 0.5$.

$b_x=1,\ b_y=1,\ b_u=1,\ b_v=1,\ \rho=0.5.$							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.002	-0.007	0.01	-0.003	1.049	1.101	0.474
E_LS	0.002	-0.007	0.009	-0.005	0.986	1.011	0.485
E_WLS	0.002	-0.007	0.009	-0.005	0.986	1.011	0.485
V_MML	0.153	0.152	0.154	0.155	0.021	0.019	0.016
V_LS	0.166	0.164	0.164	0.166	0.02	0.018	0.017
V_WLS	0.166	0.164	0.164	0.166	0.02	0.018	0.017
RE (MML/LS)	92	93	94	93	102	108	91
RE (MML/WLS)	92	93	94	93	102	108	91
n=30							
E_MML	0.002	0.003	0	-0.01	1.032	1.063	0.483
E_LS	0.003	0.004	-0.003	-0.012	0.991	1.007	0.49
E_WLS	0.003	0.004	-0.003	-0.012	0.991	1.007	0.49
V_MML	0.102	0.103	0.103	0.097	0.013	0.012	0.011
V_LS	0.11	0.111	0.111	0.106	0.013	0.012	0.011
V_WLS	0.11	0.111	0.111	0.106	0.013	0.012	0.011
RE (MML/LS)	93	93	92	91	98	99	92
RE (MML/WLS)	93	93	92	91	98	99	92
n=100							
E_MML	-0.001	0.004	-0.003	-0.008	1.007	1.016	0.493
E_LS	-0.003	0.002	-0.003	-0.009	0.995	1	0.495
E_WLS	-0.003	0.002	-0.003	-0.009	0.995	1	0.495
V_MML	0.03	0.029	0.033	0.03	0.004	0.003	0.003
V_LS	0.033	0.03	0.035	0.033	0.004	0.003	0.004
V_WLS	0.033	0.03	0.035	0.033	0.004	0.003	0.004
RE (MML/LS)	93	97	92	90	90	93	91
RE (MML/WLS)	93	97	92	90	90	93	91

Table 3.1.8 Simulation results (GL), b_x=0.5, b_y=1, b_u=1, b_v=0.5, $\rho = 0.5$.

b_x=0.5, b_y=1, b_u=1, b_v=0.5, $\rho = 0.5$.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	-0.005	-0.01	0.01	0.004	1.048	1.098	0.477
E_LS	-0.02	-0.018	0.009	-0.005	0.982	1.011	0.485
E_WLS	-0.025	-0.02	0.009	-0.006	0.979	1.009	0.483
V_MML	0.258	0.211	0.154	0.235	0.021	0.018	0.014
V_LS	0.284	0.232	0.164	0.258	0.027	0.022	0.018
V_WLS	0.277	0.226	0.164	0.271	0.033	0.02	0.017
RE (MML/LS)	91	91	94	91	79	83	78
RE (MML/WLS)	93	93	94	87	64	94	84
n=30							
E_MML	-0.002	0.001	0	-0.007	1.031	1.062	0.485
E_LS	-0.012	-0.003	-0.003	-0.014	0.987	1.006	0.49
E_WLS	-0.016	-0.004	-0.003	-0.014	0.983	1.005	0.489
V_MML	0.174	0.142	0.103	0.146	0.014	0.011	0.009
V_LS	0.191	0.155	0.111	0.164	0.018	0.015	0.012
V_WLS	0.188	0.152	0.111	0.172	0.023	0.013	0.011
RE (MML/LS)	91	91	92	89	74	75	77
RE (MML/WLS)	93	93	92	85	60	86	81
n=100							
E_MML	-0.005	0.001	-0.003	-0.008	1.006	1.016	0.494
E_LS	-0.009	-0.003	-0.003	-0.011	0.996	1.001	0.495
E_WLS	-0.01	-0.002	-0.003	-0.011	0.995	1.001	0.496
V_MML	0.051	0.038	0.033	0.044	0.004	0.003	0.003
V_LS	0.057	0.041	0.035	0.051	0.006	0.004	0.004
V_WLS	0.056	0.041	0.035	0.053	0.007	0.004	0.004
RE (MML/LS)	90	94	92	88	65	70	79
RE (MML/WLS)	91	94	92	83	51	81	79

It can be seen from these tables that the MML estimators are enormously more efficient than both the LS and the WLS estimators. A disconcerting feature of the latter is that their efficiencies relative to the MML estimators generally decrease as the sample size n increases. The cases where ρ = 0.2 and 0.9 are similar and are, therefore, not reported here.

3.1.2 Robustness

Another characteristic we would like our estimator to have is robustness. This means that it maintains its good properties under deviations from the assumed distribution and also under situations where outliers and other anomalies are present in the data. In order to study robustness properties of our MML estimators we propose some deviations from our assumed distribution and see how the estimators behave in each case. We choose the distributions most favorable to the least square method and that is the logistic distributions (b_x=1, b_y=1, b_u=1, b_v=1), and we consider three types of deviations from the assumed distribution:

1. Outlier Model:

a. Outliers among the x_i's and u_i's: Here *n-r* of the x_i's (we don't know which) come from GL(μ_x, σ_1) and *r* come from GL(μ_x, $4\sigma_1$) where r = [0.5 + 0.1 n]. Also, *n-r* of the u_i's come from GL(μ_u, σ_1) and *r* come from GL(μ_u, $4\sigma_1$) .

b. Outliers in the errors: Here *n-r* of the e_{1i}'s (we don't know which) come from GL(0, $\sigma_{2.1}$) and *r* come from GL(0, $4\sigma_{2.1}$) where r = [0.5 + 0.1 n]. The same holds for the e_{2i}'s.

c. Outliers among the x_i's , u_i's and also in the errors: This is a combination of cases (a) and (b) above.

2. Mixture Model:

a. A mixture in the x_i's and u_i's:

- 0.90 GL(μ_x, σ_1) + 0.10 GL(μ_x, 4σ_1): 90% of the x_i observations come from GL(μ_x, σ_1) and 10% come from GL(μ_x, 4σ_1).
- 0.90 GL(μ_u, σ_1) + 0.10 GL(μ_u, 4σ_1): 90% of the u_i's come from GL(μ_u, σ_1) and 10% come from GL(μ_u, 4σ_1).

b. A mixture among the errors: 0.90 GL(0, $\sigma_{2.1}$) + 0.10 GL(0, 4$\sigma_{2.1}$). 90% of the e_{1i}'s come from GL(0, $\sigma_{2.1}$) and 10% come from GL(0, 4$\sigma_{2.1}$). The same holds for the e_{2i}'s.

c. A mixture among the x_i's , u_i's and also in the errors: This is a combination of cases (a) and (b) above.

3. Contamination Model:

a. Contamination in the x_i's and u_i's:

- 0.90 GL(μ_x, σ_1) + 0.10 U(a_1,b_1). 90% of the x_i's come from GL(μ_x, σ_1) and 10% come from a uniform distribution U(a_1,b_1) where, a_1= μ_x- $\sigma_1/2$ and b_1= μ_x+ $\sigma_1/2$. Here since we assume μ_x= 0 and σ_1=1, we have a_1= - 0.5 and b_1= 0.5.
- 0.90 GL(μ_u, σ_1) + 0.10 U(a_2,b_2). 90% of the u_i's come from GL(μ_x, σ_1) and 10% come from a uniform distribution U(a_2,b_2) where, a_2= μ_u- $\sigma_1/2$ and b_2= μ_u+ $\sigma_1/2$. Since we assume μ_u= 0 and σ_1=1, we have a_2= - 0.5 and b_2= 0.5.

b. Contamination in the errors: 0.90 GL(0, $\sigma_{2.1}$) + 0.10 U(a, b). 90% of the e_{1i}'s come from GL(0, $\sigma_{2.1}$) and 10% come from U(a, b), where a = -$\sigma_{2.1}$/2 and b = $\sigma_{2.1}$/2. The same holds for the e_{2i}'s.

c. Contamination in the x_i's and u_i's and also in the errors: This is a combination of cases (a) and (b) above.

We consider the above models with 3 different values of ρ (0.5, 0.2 and 0.9). It may be noted that in the models above Var(X), Var(U) and the variances of the errors change from the usual cases where no deviation from the model is assumed. Thus, the true values of σ_1, σ_2 and ρ are shifted. In order to compare the MML and the LS estimators below we can compare their variances to see which has a lower variance. Thus, we will simply look at the RE (relative efficiency) of the LSE to the MMLE. Since we consider only the case where b_x=1, b_y=1, b_u=1, b_v=1, the LSE and the weighted LSE are exactly the same. Therefore, we only report the relative efficiency with respect to the LS estimators. Although we have not reported the simulated means, but the means of the LS and the MML estimators are essentially the same. In fact, the bias in all the estimators is negligible. Notice from the tables below how inefficient the LS estimators are relative to the MML estimators, even for small n (n = 20). The LS estimators are highly affected if there are outliers in the data. The MML estimators are calculated such that small weights are assigned to extreme observations. Thus, they are less affected when there are outliers in the data.

1. Outlier Model

Table 3.1.9 Simulation results, outliers among the x_i's and u_i's (GL).

ρ=0.5, Outlier Model (a), b_x=1, b_y=1, b_u=1, b_v=1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
V_MML	0.228	0.172	0.229	0.172	0.094	0.027	0.014
RE (MML/LS)	57	78	56	77	58	72	83
n=30							
V_MML	0.148	0.117	0.149	0.108	0.054	0.016	0.008
RE (MML/LS)	54	75	53	73	46	62	78
n=100							
V_MML	0.042	0.034	0.04	0.034	0.012	0.004	0.002
RE (MML/LS)	51	76	54	76	30	49	68

Table 3.1.10 Simulation results, outliers among the errors (GL).

ρ=0.5, Outlier Model (b), b_x=1, b_y=1, b_u=1, b_v=1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
V_MML	0.153	0.209	0.153	0.211	0.02	0.072	0.017
RE (MML/LS)	92	62	93	62	104	60	71
n=30							
V_MML	0.102	0.139	0.103	0.128	0.013	0.042	0.011
RE (MML/LS)	93	59	93	56	97	48	67
n=100							
V_MML	0.03	0.04	0.03	0.04	0.004	0.009	0.003
RE (MML/LS)	93	56	96	55	89	32	65

Table 3.1.11 Simulation results, outliers among the errors, the x_i's and the u_i's (GL).

ρ =0.5, Outlier Model (c), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
V_MML	0.228	0.228	0.229	0.231	0.094	0.09	0.037
RE (MML/LS)	57	58	56	58	58	60	66
n=30							
V_MML	0.148	0.152	0.149	0.14	0.054	0.053	0.024
RE (MML/LS)	54	55	53	52	46	49	60
n=100							
V_MML	0.042	0.043	0.04	0.043	0.012	0.011	0.007
RE (MML/LS)	51	51	54	51	30	34	53

2. Mixture Model

Table 3.1.12 Simulation results, mixture in the x_i's and u_i's (GL).

ρ =0.5, Mixture Model (a), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
V_MML	0.233	0.175	0.241	0.171	0.127	0.029	0.016
RE (MML/LS)	58	79	58	77	64	74	84
n=30							
V_MML	0.151	0.114	0.149	0.113	0.074	0.018	0.011
RE (MML/LS)	55	76	54	75	53	64	80
n=100							
V_MML	0.042	0.033	0.04	0.03	0.016	0.004	0.002
RE (MML/LS)	52	75	49	71	35	50	69

Table 3.1.13 Simulation results, Mixture among the errors (GL).

ρ =0.5, Mixture Model (b), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
V_MML	0.153	0.21	0.155	0.217	0.02	0.096	0.018
RE (MML/LS)	93	63	93	61	102	67	73
n=30							
V_MML	0.1	0.136	0.099	0.135	0.013	0.053	0.012
RE (MML/LS)	92	59	91	60	97	53	70
n=100							
V_MML	0.033	0.04	0.029	0.036	0.003	0.013	0.004
RE (MML/LS)	91	56	94	55	87	36	64

Table 3.1.14 Simulation results, mixture among the errors, the x_i's and the u_i's (GL).

ρ =0.5, Mixture Model (c), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
V_MML	0.227	0.243	0.237	0.244	0.125	0.1	0.021
RE (MML/LS)	57	58	57	58	63	68	70
n=30							
V_MML	0.145	0.145	0.146	0.152	0.073	0.054	0.014
RE (MML/LS)	54	56	53	55	52	53	63
n=100							
V_MML	0.039	0.043	0.04	0.044	0.018	0.011	0.004
RE (MML/LS)	48	52	50	50	37	36	51

3.Contamination Model:

Table 3.1.15 Simulation results, contamination in the x_i 's and u_i 's (GL).

ρ =0.5, Contamination Model (a), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
V_MML	0.127	0.142	0.127	0.144	0.02	0.019	0.016
RE (MML/LS)	87	92	86	91	100	108	91
n=30							
V_MML	0.086	0.1	0.083	0.095	0.013	0.012	0.011
RE (MML/LS)	85	90	86	91	96	99	90
n=100							
V_MML	0.024	0.028	0.024	0.028	0.004	0.003	0.003
RE (MML/LS)	82	90	84	95	91	91	92

Table 3.1.16 Simulation results, contamination among the errors (GL).

ρ =0.5, Contamination Model (b), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
V_MML	0.153	0.133	0.146	0.131	0.021	0.019	0.014
RE (MML/LS)	93	89	93	89	103	105	88
n=30							
V_MML	0.104	0.089	0.097	0.089	0.013	0.012	0.009
RE (MML/LS)	91	88	92	89	96	98	87
n=100							
V_MML	0.031	0.027	0.03	0.025	0.004	0.003	0.003
RE (MML/LS)	91	85	93	86	92	92	86

Table 3.1.17 Simulation results, contaminations among the errors, the x_i's and the u_i's (GL).

ρ =0.5, Contamination Model (c), b_x=1, b_y=1, b_u=1, b_v=1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
V_MML	0.126	0.126	0.133	0.127	0.02	0.018	0.016
RE (MML/LS)	86	88	87	88	103	105	89
n=30							
V_MML	0.086	0.087	0.082	0.083	0.013	0.011	0.01
RE (MML/LS)	85	86	85	86	96	97	88
n=100							
V_MML	0.024	0.025	0.025	0.026	0.004	0.003	0.003
RE (MML/LS)	84	86	83	82	91	90	85

The results for ρ = 0.2 and 0.9 are found in appendix D.

3.2 Student's t distribution

Here, we also consider different values of the degrees of freedom for each t distribution. We repeat the above simulation study for the t-distribution.

3.2.1 Efficiency

Here, we compare the variances of our MMLE to the corresponding diagonal elements of the inverse of Fisher information matrix. We notice how close the values are, even for small n. Note that we have multiplied both entries by n.

Table 3.2.1 Simulated variances of the MMLE vs. the diagonal elements of I^{-1} (Student's t), $r_x = r_y = r_u = r_v = 4$.

n=20	μ_x	μ_y	μ_u	μ_v	$\mu_{y/x}$	$\mu_{v/u}$	σ_1	σ_2	$\sigma_{2.1}$	ρ
n*V_MML	1.54	1.61	1.52	1.52	1.20	1.18	1.02	0.94	0.90	0.34
n*Fisher	1.40	1.40	1.40	1.40	1.05	1.05	0.44	0.34	0.33	0.27
n=30										
n*V_MML	1.49	1.47	1.49	1.46	1.11	1.13	0.82	0.67	0.67	0.33
n*Fisher	1.40	1.40	1.40	1.40	1.05	1.05	0.44	0.34	0.33	0.27
n=50										
n*V_MML	1.535	1.50	1.46	1.38	1.10	1.08	0.78	0.58	0.56	0.33
n*Fisher	1.4	1.40	1.40	1.40	1.05	1.05	0.44	0.34	0.33	0.27
n=100										
n*V_MML	1.50	1.53	1.33	1.34	1.11	1.01	0.58	0.45	0.43	0.30
n*Fisher	1.40	1.40	1.40	1.40	1.05	1.05	0.44	0.34	0.33	0.27

Similar tables can be obtained for different degrees of freedom.

We now compare our MML estimators to the LS estimators and consider three values of the degrees of freedom, as shown in the tables that follow.

Table 3.2.2 Simulation results (Student's t), $r_x = r_y = r_u = r_v = 4$, $\rho = 0.5$.

$r_x=4$, $r_y=4$, $r_u=4$, $r_v=4$, $\rho=0.5$.							
n = 20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	-0.004	-0.002	0.004	0.001	1.157	1.212	0.478
E_LS	-0.003	-0.003	0.005	0.001	0.976	1	0.489
E_WLS	-0.003	-0.003	0.005	0.001	0.976	1	0.489
V_MML	0.077	0.081	0.076	0.076	0.051	0.047	0.017
V_LS	0.103	0.105	0.099	0.1	0.054	0.047	0.022
V_WLS	0.103	0.105	0.099	0.1	0.054	0.047	0.022
RE (MML/LS)	75	76	76	76	95	101	79
RE (MML/WLS)	75	76	76	76	95	101	79
n = 30							
E_MML	0.005	0.01	0.003	0	1.11	1.145	0.485
E_LS	0.009	0.014	0.007	0.003	0.979	0.999	0.491
E_WLS	0.009	0.014	0.007	0.003	0.979	0.999	0.491
V_MML	0.05	0.049	0.05	0.049	0.027	0.022	0.011
V_LS	0.067	0.066	0.066	0.066	0.036	0.028	0.015
V_WLS	0.067	0.066	0.066	0.066	0.036	0.028	0.015
RE (MML/LS)	74	75	75	74	75	79	74
RE (MML/WLS)	74	75	75	74	75	79	74
n=100							
E_MML	-0.003	-0.004	-0.001	0	1.043	1.053	0.493
E_LS	-0.003	-0.004	-0.001	0.002	0.987	0.992	0.494
E_WLS	-0.003	-0.004	-0.001	0.002	0.987	0.992	0.494
V_MML	0.015	0.015	0.013	0.013	0.006	0.005	0.003
V_LS	0.02	0.022	0.019	0.019	0.011	0.009	0.005
V_WLS	0.02	0.022	0.019	0.019	0.011	0.009	0.005
RE (MML/LS)	74	71	69	71	51	53	56
RE (MML/WLS)	74	71	69	71	51	53	56

Table 3.2.3 Simulation results (Student's t), r_x=4, r_y=6, r_u=4, r_v=6, ρ=0.5.

r_x=4, r_y=6, r_u=4, r_v=6, ρ=0.5.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.009	0.004	0.003	0.007	1.154	1.143	0.501
E_LS	0.01	0.004	0.001	0.007	0.972	1.005	0.48
E_WLS	0.01	0.004	0.001	0.007	0.972	1.005	0.48
V_MML	0.075	0.068	0.074	0.068	0.05	0.024	0.015
V_LS	0.101	0.08	0.097	0.08	0.053	0.023	0.018
V_WLS	0.101	0.08	0.097	0.08	0.053	0.023	0.018
RE (MML/LS)	75	86	76	85	95	101	88
RE (MML/WLS)	75	86	76	85	95	101	88
n=30							
E_MML	-0.006	0.001	0.006	0	1.114	1.097	0.506
E_LS	-0.003	0.002	0.007	-0.002	0.984	1.006	0.487
E_WLS	-0.003	0.002	0.007	-0.002	0.984	1.006	0.487
V_MML	0.05	0.045	0.05	0.045	0.046	0.017	0.01
V_LS	0.073	0.056	0.068	0.053	0.073	0.024	0.012
V_WLS	0.073	0.056	0.068	0.053	0.073	0.024	0.012
RE (MML/LS)	69	81	74	85	63	72	83
RE (MML/WLS)	69	81	74	85	63	72	83
n=100							
E_MML	-0.005	-0.003	0.002	-0.002	1.048	1.043	0.5
E_LS	-0.006	-0.004	0.003	0	0.99	1.001	0.491
E_WLS	-0.006	-0.004	0.003	0	0.99	1.001	0.491
V_MML	0.014	0.013	0.015	0.013	0.006	0.004	0.003
V_LS	0.02	0.015	0.021	0.016	0.012	0.005	0.004
V_WLS	0.02	0.015	0.021	0.016	0.012	0.005	0.004
RE (MML/LS)	72	85	73	80	49	74	68
RE (MML/WLS)	72	85	73	80	49	74	68

Table 3.2.4 Simulation results (Student's t), r_x =6, r_y =3, r_u =3, r_v =6, ρ =0.5.

r_x =6, r_y =3, r_u =3, r_v =6, ρ =0.5.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	-0.002	-0.002	0.000	-0.001	1.145	1.192	0.481
E_LS	-0.003	-0.002	0.003	0.000	0.948	0.966	0.492
E_WLS	-0.003	-0.002	0.003	0.000	0.934	0.973	0.477
V_MML	0.066	0.08	0.084	0.070	0.057	0.065	0.018
V_LS	0.076	0.127	0.147	0.092	0.086	0.077	0.025
V_WLS	0.076	0.127	0.147	0.092	0.109	0.063	0.024
RE (MML/LS)	87	63	57	76	67	84	72
RE (MML/WLS)	87	63	57	76	53	102	74
n=30							
E_MML	0.000	-0.001	0.002	0.004	1.112	1.140	0.490
E_LS	-0.002	0.000	0.003	0.004	0.963	0.977	0.497
E_WLS	-0.002	0.000	0.003	0.004	0.952	0.983	0.485
V_MML	0.042	0.051	0.056	0.047	0.031	0.031	0.011
V_LS	0.049	0.088	0.102	0.064	0.063	0.071	0.018
V_WLS	0.049	0.088	0.102	0.064	0.081	0.056	0.018
RE (MML/LS)	86	59	54	72	50	43	64
RE (MML/WLS)	86	59	54	72	39	55	64
n=100							
E_MML	-0.001	-0.006	0.000	0.005	1.048	1.051	0.500
E_LS	0.000	-0.006	0.002	0.005	0.976	0.975	0.504
E_WLS	0.000	-0.006	0.002	0.005	0.969	0.981	0.495
V_MML	0.012	0.015	0.015	0.013	0.007	0.004	0.003
V_LS	0.014	0.024	0.029	0.018	0.059	0.023	0.007
V_WLS	0.014	0.024	0.029	0.018	0.074	0.021	0.007
RE (MML/LS)	85	63	53	74	12	19	43
RE (MML/WLS)	85	63	53	74	10	21	44

3.2.2 Robustness

We choose the case of equal degrees of freedom for each variable (r_x = 4, r_y = 4, r_u = 4, r_v = 4), and we will assume the following deviations from the assumed distribution:

Outlier Model:

a. Outliers among the x_i's and u_i's: Here *n-r* of the x_i's (we don't know which) come from the distribution with p.d.f. given in equation (2.5.1); and *r* of the x_i's come from same distribution with σ_1 multiplied by 4. Also *n-r* of the u_i's (we don't know which) come from the distribution with p.d.f. given in equation (2.5.3); and *r* of the u_i's come from same distribution with σ_1 multiplied by 4, where r = [0.5 + 0.1 n].
b. Outliers in the errors: Here *n-r* of the e_{1i}'s (we don't know which) come from a distribution with p.d.f. given in equation (2.5.2) and *r* come from the same distribution with $\sigma_{2.1}$ multiplied by 4. n - r of the e_{2i}'s (we don't know which) come from a distribution with p.d.f. given in equation (2.5.4) and *r* come from the same distribution $\sigma_{2.1}$ multiplied by 4, where r = [0.5 + 0.1 n].
c. Outliers among the x_i's , u_i's and also in the errors: This is a combination of cases (a) and (b) above.

Note again that the true values of σ_1, σ_2 and ρ are shifted. In order to compare the MML and the LS estimators below we simply compare their variances to see which has a lower variance. Thus, we will simply look at the

RE (relative efficiency) of the LSE to the MMLE. Since we consider only the case where $r_x = 4$, $r_y = 4$, $r_u = 4$, $r_v = 4$, the LSE and the weighted LSE are exactly the same and thus we only compare the LS to the MML estimators.

Table 3.2.5 Simulation results, outliers among the x_i's and u_i's (Student's t).

Outlier Model (a), $r_x = 4$, $r_y = 4$, $r_u = 4$, $r_v = 4$							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.005	0.005	-0.003	0.000	1.639	1.351	0.601
E_LS	0.004	0.005	-0.001	0.000	1.485	1.156	0.631
V_MML	0.116	0.086	0.113	0.087	0.224	0.061	0.018
RE (MML/LS)	46	63	46	62	78	84	78
n=30							
E_MML	-0.004	-0.003	-0.006	0.002	1.551	1.27	0.605
E_LS	-0.008	-0.003	0.001	0.009	1.511	1.161	0.641
V_MML	0.069	0.054	0.069	0.054	0.135	0.037	0.012
RE (MML/LS)	40	59	41	59	55	62	72
n=100							
E_MML	0.003	0.005	-0.004	0.002	1.412	1.154	0.609
E_LS	0.003	0.005	-0.003	0.000	1.569	1.169	0.664
V_MML	0.02	0.016	0.019	0.016	0.054	0.012	0.003
RE (MML/LS)	34	56	38	55	17	17	53

Table 3.2.6 Simulation results, outliers among the errors (Student's t).

Outlier Model (b), $r_x = 4,\ r_y = 4,\ r_u = 4,\ r_v = 4$							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.007	0.012	-0.004	-0.002	1.156	1.62	0.368
E_LS	0.005	0.009	-0.007	-0.005	0.975	1.415	0.36
V_MML	0.076	0.108	0.073	0.107	0.049	0.188	0.02
RE (MML/LS)	74	51	76	50	96	87	72
n=30							
E_MML	-0.004	-0.001	0.001	0.000	1.109	1.502	0.378
E_LS	0.000	-0.003	0.001	-0.001	0.98	1.422	0.359
V_MML	0.048	0.066	0.051	0.066	0.027	0.104	0.013
RE (MML/LS)	71	44	76	46	76	59	67
n=100							
E_MML	0.000	-0.004	0.002	0.001	1.049	1.336	0.397
E_LS	-0.002	-0.006	0.001	0.004	0.995	1.437	0.357
V_MML	0.016	0.018	0.014	0.020	0.006	0.015	0.003
RE (MML/LS)	71	45	74	45	40	23	46

Table 3.2.7 Simulation results, outliers among the errors, the x_i's and u_i's (Student's t).

ρ=0.5, Outlier Model (c), $r_x = 4$, $r_y = 4$, $r_u = 4$, $r_v = 4$							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.003	0.005	-0.002	0.002	1.633	1.715	0.476
E_LS	0.001	0.004	-0.004	-0.001	1.478	1.517	0.484
V_MML	0.113	0.114	0.115	0.115	0.206	0.206	0.04
RE (MML/LS)	48	48	45	46	79	85	69
n=30							
E_MML	0.004	0.006	0.002	0.000	1.544	1.604	0.478
E_LS	0.005	0.007	0.007	0.001	1.502	1.539	0.484
V_MML	0.072	0.072	0.068	0.071	0.12	0.133	0.027
RE (MML/LS)	43	42	41	44	56	70	63
n=100							
E_MML	0.004	0.002	0.000	-0.007	1.394	1.417	0.489
E_LS	0.006	-0.001	-0.007	-0.015	1.529	1.556	0.488
V_MML	0.019	0.021	0.019	0.02	0.022	0.016	0.006
RE (MML/LS)	40	39	39	38	23	25	38

Simulations were done with mixture and contamination models as well. The MMLE were found to be more robust than the LSE having lower variance and less bias in all the cases. We do not reproduce the results as they are similar to those given in Tables 3.1.12-3.1.17.

Fortran programs used in the simulations are given in Appendix E.

CHAPTER 4

HYPOTHESIS TESTING

We are often interested in testing the hypothesis that two population means are the same. Thus we may be interested in testing a hypothesis such as:

$$H_0 : \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} \mu_u \\ \mu_v \end{bmatrix}.$$

Note that since $\mu_y = \mu_{y/x} + \theta\, \mu_x$ and $\mu_v = \mu_{v/u} + \theta\, \mu_u$, testing the hypothesis H_0 above is equivalent to testing:

$$H_0 : \begin{bmatrix} \mu_x \\ \mu_{y/x} \end{bmatrix} = \begin{bmatrix} \mu_u \\ \mu_{v/u} \end{bmatrix} \text{ or equivalently, } H_0 : \begin{bmatrix} \mu_x - \mu_u \\ \mu_{y/x} - \mu_{v/u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

In order to test H_0, we define a Hotelling type T^2 statistic based on the MMLE given in sections 2.4 and 2.5. We denote it by $\hat{T}^2$. We also give the corresponding test based on the LSE, $\tilde{T}^2$. We derive the noncentrality parameters in the asymptotic distributions of $\hat{T}^2$ and $\tilde{T}^2$ and, thereby, show that the former has higher power. For small to moderate sample sizes, we simulate values of the power. We show that the $\hat{T}^2$ test has higher power than the $\tilde{T}^2$ test. We compare $\hat{T}^2$ with the test statistic given by Tiku and Singh (1982), T_D^2, which is based on censored samples. We show that, for testing H_0, $\hat{T}^2$ has overall higher power than the T_D^2 test. It is interesting to note how well T_D^2 competes with the sophisticated test statistic, $\hat{T}^2$, for symmetric bivariate distributions. Finally, we show that using the MMLE derived in sections 2.4 and 2.5, we can obtain an efficient estimator of Mahalanobis distance. We show that, using the MMLE, our estimator of Mahalanobis distance has less bias and smaller variance than if we were to use the LSE.

4.1 Generalized Logistic Distribution

4.1.1 The test based on the MML estimators

Lemma 4.1.1

(a) The MMLE $\hat{\mu}_x$ (for a given σ_1)is the BAN (best asymptotically normal) estimator of μ_x and has asymptotic variance $Var(\hat{\mu}_x) \cong \dfrac{\sigma_1^2}{(b_x+1)m_{11}}$.

(b) The MMLE $\hat{\mu}_u$ (for a given σ_1) is the BAN estimator of μ_u and has asymptotic variance $Var(\hat{\mu}_u) \cong \dfrac{\sigma_1^2}{(b_u+1)m_{21}}$.

Proof:

We can write

$$\frac{\partial \ln L^*}{\partial \mu_x} = \frac{n}{\sigma_1} - \frac{(b_x+1)}{\sigma_1}\sum_{i=1}^{n}(\alpha_{1i} - \beta_{1i} z_{1(i)})$$

$$= \frac{(b_x+1)}{\sigma_1}\left[\sum_{i=1}^{n}\left((b_x+1)^{-1} - \alpha_{1i}\right) + \frac{1}{\sigma_1}\sum_{i=1}^{n}\beta_{1i}(x_{(i)} - \mu_x)\right]$$

$$= \frac{(b_x+1)}{\sigma_1}\left[\frac{1}{\sigma_1}\sum_{i=1}^{n}\beta_{1i}x_{(i)} - D_{11}m_{11} - \frac{\mu_x}{\sigma_1}m_{11}\right]$$

$$= \frac{(b_x+1)m_{11}}{\sigma_1^2}\left[K_{11} - D_{11}\sigma_1 - \mu_x\right].$$

Thus,

$$\frac{\partial \ln L^*}{\partial \mu_x} = \frac{(b_x+1)m_{11}}{\sigma_1^2}\left[\hat{\mu}_x(\sigma_1) - \mu_x\right].$$

Since $\lim_{n\to\infty}\frac{1}{n}\left\{\frac{\partial \ln L}{\partial \mu_x} - \frac{\partial \ln L^*}{\partial \mu_x}\right\} = 0$,

the modified maximum likelihood equation $\frac{\partial \ln L^*}{\partial \mu_x} = 0$ is asymptotically equivalent to the maximum likelihood equation $\frac{\partial \ln L}{\partial \mu_x} = 0$. From the above form of $\frac{\partial \ln L^*}{\partial \mu_x}$, the result follows.

The same is true for $\hat{\mu}_u$ since

$$\frac{\partial \ln L}{\partial \mu_u} \cong \frac{\partial \ln L^*}{\partial \mu_u} = \frac{(b_u+1)m_{21}}{\sigma_1^2}\left[\hat{\mu}_u(\sigma_1) - \mu_u\right].$$

Therefore, $\hat{\mu}_u$ (for a given σ_1) is the BAN estimator with asymptotic variance $Var(\hat{\mu}_u) \cong \frac{\sigma_1^2}{(b_u+1)m_{21}}$.

Lemma 4.1.2:

(a) The MMLE $\hat{\mu}_{y/x}$ (for given θ and $\sigma_{2.1}$) is the BAN estimator of $\mu_{y/x}$ and has asymptotic variance $Var(\hat{\mu}_{y/x}) \cong \dfrac{\sigma_{2.1}^2}{(b_y+1)m_{12}}$.

(b) The MMLE $\hat{\mu}_{v/u}$ (for given θ and $\sigma_{2.1}$) is the BAN estimator of $\mu_{v/u}$ and has asymptotic variance $Var(\hat{\mu}_{v/u}) \cong \dfrac{\sigma_{2.1}^2}{(b_v+1)m_{22}}$.

Proof: $\dfrac{\partial \ln L^*}{\partial \mu_{y/x}}$ can be written as

$$\frac{\partial \ln L^*}{\partial \mu_{y/x}} = \frac{n}{\sigma_{2.1}} - \frac{(b_y+1)}{\sigma_{2.1}}\sum_{i=1}^{n}(\alpha_{2i} - \beta_{2i} z_{2(i)})$$

$$= \frac{(b_y+1)}{\sigma_{2.1}^2}\left[-\Delta_1 + \frac{1}{\sigma_{2.1}}\left(\sum_{i=1}^{n}\beta_{2i} y_{(i)} - \mu_{y/x}\sum_{i=1}^{n}\beta_{2i} - \theta\sum_{i=1}^{n}\beta_{2i} x_{(i)}\right)\right]$$

$$= \frac{(b_y+1)m_{12}}{\sigma_{2.1}^2}\left[\bar{y}_{[.]} - \theta\bar{x}_{[.]} - \frac{\Delta_1\sigma_{2.1}}{m_{12}} - \mu_{y/x}\right].$$

Thus, we have (for given θ and $\sigma_{2.1}$)

$$\frac{\partial \ln L}{\partial \mu_{y/x}} \cong \frac{\partial \ln L^*}{\partial \mu_{y/x}} = \frac{(b_y+1)m_{12}}{\sigma_{2.1}^2}\left[\hat{\mu}_{y/x} - \mu_{y/x}\right].$$

Therefore, the conditional distribution of $\hat{\mu}_{y/x}$ (given θ and $\sigma_{2.1}$) is asymptotically normal with mean $\mu_{y/x}$ and asymptotic variance

$Var(\hat{\mu}_{y/x}) \cong \dfrac{\sigma_{2.1}^2}{(b_y+1)m_{12}}$. Since $\hat{\theta}$ converges to θ as n tends to infinity, the conditionality on θ can be removed.

Similarly, we can write

$$\frac{\partial \ln L}{\partial \mu_{v/u}} \cong \frac{\partial \ln L^*}{\partial \mu_{v/u}} = \frac{(b_v+1)m_{22}}{\sigma_{2.1}^2}\left[\hat{\mu}_{v/u} - \mu_{v/u}\right].$$

Therefore, $\hat{\mu}_{v/u}$ (for given θ and $\sigma_{2.1}$) is BAN with asymptotic variance $Var(\hat{\mu}_{v/u}) \cong \dfrac{\sigma_{2.1}^2}{(b_v+1)m_{22}}$.

We have shown that $\hat{\mu}_x$ and $\hat{\mu}_{y/x}$ are asymptotically normal. Thus, the distribution of the vector $\sqrt{n}\left(\hat{\mu}_x, \hat{\mu}_{y/x}\right)$ is asymptotically bivariate normal with mean vector $\sqrt{n}\left(\mu_x, \mu_{y/x}\right)$ and estimated variance-covariance matrix

$$\hat{\Omega}_1 = \begin{bmatrix} \hat{\sigma}_{11} & 0 \\ 0 & \hat{\sigma}_{44} \end{bmatrix}, \text{ where } \hat{\sigma}_{11} = \frac{n\hat{\sigma}_1^2}{m_{11}(b_x+1)} \text{ and } \hat{\sigma}_{44} = \frac{n\hat{\sigma}_{2.1}^2}{m_{12}(b_y+1)}.$$

Also since $\hat{\mu}_u$ and $\hat{\mu}_{v/u}$ are asymptotically normal, the distribution of the vector $\sqrt{n}\left(\hat{\mu}_u, \hat{\mu}_{v/u}\right)$ is asymptotically bivariate normal with mean vector $\sqrt{n}\left(\mu_u, \mu_{v/u}\right)$ and estimated variance-covariance matrix

$$\hat{\Omega}_2 = \begin{bmatrix} \hat{\sigma}_{22} & 0 \\ 0 & \hat{\sigma}_{55} \end{bmatrix}, \text{ where } \hat{\sigma}_{22} = \frac{n\hat{\sigma}_1^2}{m_{21}(b_u+1)}, \hat{\sigma}_{55} = \frac{n\hat{\sigma}_{2.1}^2}{m_{22}(b_v+1)}.$$

It follows that the distribution of the vector $\sqrt{n}\left(\hat{\mu}_x - \hat{\mu}_u, \hat{\mu}_{y/x} - \hat{\mu}_{v/u}\right)$ under H_0 is asymptotically bivariate normal with zero mean vector and an estimated variance-covariance matrix

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_{xu}^2 & 0 \\ 0 & \hat{\sigma}_{yv}^2 \end{bmatrix}, \text{ where}$$

$$\hat{\sigma}_{xu}^2 = n\hat{\sigma}_1^2 \left[\frac{1}{m_{11}(b_x+1)} + \frac{1}{m_{21}(b_u+1)} \right] \text{ and}$$

$$\hat{\sigma}_{yv}^2 = n\hat{\sigma}_{2.1}^2 \left[\frac{1}{m_{12}(b_y+1)} + \frac{1}{m_{22}(b_v+1)} \right].$$

Thus, to test H_0 we define the test statistic:

$$\hat{T}^2 = n\left[\hat{\mu}_x - \hat{\mu}_u \quad \hat{\mu}_{y/x} - \hat{\mu}_{v/u}\right] \hat{\Omega}^{-1} \begin{bmatrix} \hat{\mu}_x - \hat{\mu}_u \\ \hat{\mu}_{y/x} - \hat{\mu}_{v/u} \end{bmatrix}; \tag{4.1.1}$$

which can also be written as

$$\hat{T}^2 = \frac{n\left(\hat{\mu}_x - \hat{\mu}_u\right)^2}{\hat{\sigma}_{xu}^2} + \frac{n\left(\hat{\mu}_{y/x} - \hat{\mu}_{v/u}\right)^2}{\hat{\sigma}_{yv}^2}.$$

Since $\hat{\sigma}_1$ converges to σ_1 and $\hat{\sigma}_{2.1}$ converges to $\sigma_{2.1}$ as n goes to infinity, the null distribution of $\hat{T}^2$ is asymptotically chi-square with 2 degrees of freedom.

Under H_1, asymptotically, the distribution of $\hat{T}^2$ is non-central chi-square with 2 degrees of freedom and non-centrality parameter

$$\tau^2 = \frac{n(\mu_x - \mu_u)^2}{\sigma_{xu1}^2} + \frac{n(\mu_{y/x} - \mu_{v/u})^2}{\sigma_{yv1}^2}, \text{ where}$$

$$\sigma_{xu1}^2 = n\sigma_1^2\left[\frac{1}{m_{11}(b_x+1)} + \frac{1}{m_{21}(b_u+1)}\right] \text{ and}$$

$$\sigma_{yv1}^2 = n\sigma_{2.1}^2\left[\frac{1}{m_{12}(b_y+1)} + \frac{1}{m_{22}(b_v+1)}\right].$$

4.1.2 The test based on the LS estimators

For given σ_1, $\tilde{\mu}_x = \bar{x} - (\psi(b_x) - \psi(1))\sigma_1$ and $\tilde{\mu}_u = \bar{u} - (\psi(b_u) - \psi(1))\sigma_1$.

Therefore,

$$Var(\tilde{\mu}_x) = Var(\bar{x}) = \frac{\sigma_1^2}{n}\left(\psi'(b_x) + \psi'(1)\right) \text{ and}$$

$$Var(\tilde{\mu}_u) = Var(\bar{u}) = \frac{\sigma_1^2}{n}\left(\psi'(b_u) + \psi'(1)\right).$$

Now, note that

$$\sigma_{2.1}^2(\psi'(b_y) + \psi'(1)) = Var(e_{1i}) = Var(Y_i - \theta\ X_i)$$

and, since e_{1i}'s are iid, we have

$$\sum_{i=1}^{n}\sigma_{2.1}^2(\psi'(b_y) + \psi'(1)) = \sum_{i=1}^{n}Var(Y_i - \theta\ X_i) = Var\left(n(\bar{y} - \theta\ \bar{x})\right).$$

Which gives

$$\frac{\sigma_{2.1}^2}{n}(\psi'(b_y)+\psi'(1)) = Var\left(\bar{y}-\theta\ \bar{x}\right). \qquad (4.1.2)$$

Now, for given $\sigma_{2.1}$ and θ,

$$\tilde{\mu}_{y/x} = \bar{y}-\theta\bar{x}-(\psi(b_y)-\psi(1))\sigma_{2.1}$$

and, therefore, by using equation (4.1.2)

$$Var(\tilde{\mu}_{y/x}) = Var\left(\bar{y}-\theta\bar{x}\right) = \frac{\sigma_{2.1}^2}{n}(\psi'(b_y)+\psi'(1)).$$

Similarly, for given $\sigma_{2.1}$ and θ,

$$Var(\tilde{\mu}_{v/u}) = Var\left(\bar{v}-\theta\bar{u}\right) = \frac{\sigma_{2.1}^2}{n}(\psi'(b_v)+\psi'(1)).$$

Thus, based on the LSE, we have the following test statistic for testing H_0:

$$\tilde{T}^2 = \frac{n\left(\tilde{\mu}_x-\tilde{\mu}_u\right)^2}{\tilde{\sigma}_{xu}^2}+\frac{n\left(\tilde{\mu}_{y/x}-\tilde{\mu}_{v/u}\right)^2}{\tilde{\sigma}_{yv}^2}, \text{ where} \qquad (4.1.3)$$

$$\tilde{\sigma}_{xu}^2 = \tilde{\sigma}_1^2\left(\psi'(b_x)+\psi'(b_u)+2\psi'(1)\right) \text{ and}$$

$$\tilde{\sigma}_{yv}^2 = \tilde{\sigma}_{2.1}^2\left(\psi'(b_y)+\psi'(b_v)+2\psi'(1)\right).$$

Asymptotically, the null distribution of $\tilde{T}^2$ is chi-square with 2 degrees of freedom. Under H_1, $\tilde{T}^2$ has a non-central chi-square distribution with 2 degrees of freedom and non-centrality parameter

$$\lambda^2 = \frac{n(\mu_x - \mu_u)^2}{\sigma_{xu2}^2} + \frac{n(\mu_{y/x} - \mu_{v/u})^2}{\sigma_{yv2}^2}, \text{ where}$$

$$\sigma_{xu2}^2 = \sigma_1^2(\psi'(b_x) + \psi'(b_u) + 2\psi'(1)) \text{ and}$$

$$\sigma_{yv2}^2 = \sigma_{2.1}^2(\psi'(b_y) + \psi'(b_v) + 2\psi'(1)).$$

Since $\tau^2 > \lambda^2$, the $\hat{T}^2$ test has higher power than the $\tilde{T}^2$ test.

For large n, the chi-square distribution gives fairly accurate approximation to the percentage points of $\hat{T}^2$ and $\tilde{T}^2$. For small sample sizes, we use simulations to find the percentage points.

4.1.3 Comparing $\hat{T}^2$ to $\tilde{T}^2$ (Generalized Logistic)

We have simulated the powers of the two tests, namely, the test based on the MMLE, $\hat{T}^2$, and the test based on the LSE, $\tilde{T}^2$. We carry out simulations for different values of n and for $b_x = b_u = b_y = b_v = 1$ (the most favorable situation for the $\tilde{T}^2$ test). Note that we have used the simulated percentage points in each case making sure that the probability of type I error (α) is 0.05 for both tests. The graphs given below are those of the power for increasing values of the noncentrality parameter, H_0 and H_1 being

$$H_0 : \begin{bmatrix} \mu_x - \mu_u \\ \mu_{y/x} - \mu_{v/u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } H_1 : \begin{bmatrix} \mu_x - \mu_u \\ \mu_{y/x} - \mu_{v/u} \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \end{bmatrix}.$$

The dotted line represents the power values of the $\tilde{T}^2$ test and the solid line represents the power values of the $\hat{T}^2$ test.

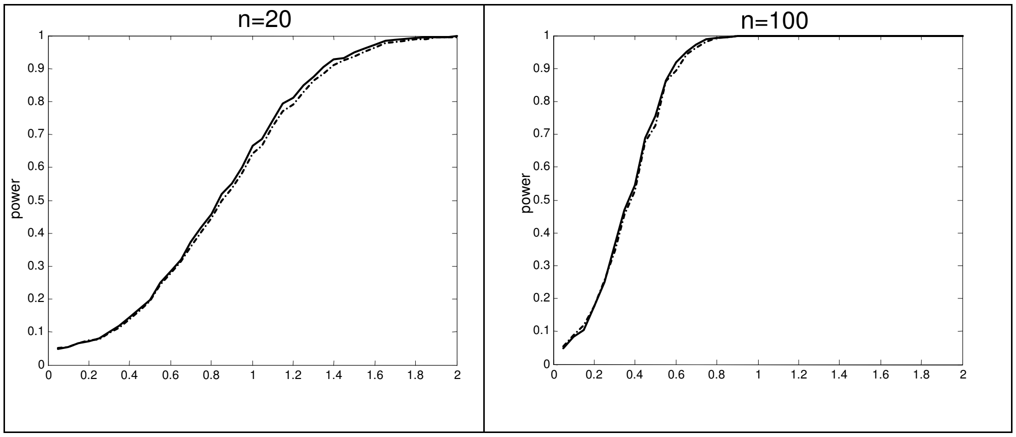

Figure 4.1.1 Power graphs of $\hat{T}^2$ and $\tilde{T}^2$ (GL), no data anomalies.

We notice that the $\hat{T}^2$ test has slightly higher power than the $\tilde{T}^2$ test. There is a small difference between them. This is because we have assumed that $b_x = 1, b_y = 1, b_u = 1, b_v = 1$ which gives an advantage to the LS method because the GL distribution with b = 1 is very close to the normal distribution. When the underlying distributions are skew the $\hat{T}^2$ test has much higher power than the $\tilde{T}^2$ test. Also, in the presence of outliers in the data or other data anomalies, notice the big difference between the two tests with $\hat{T}^2$ being superior as shown in Figure 4.1.2.

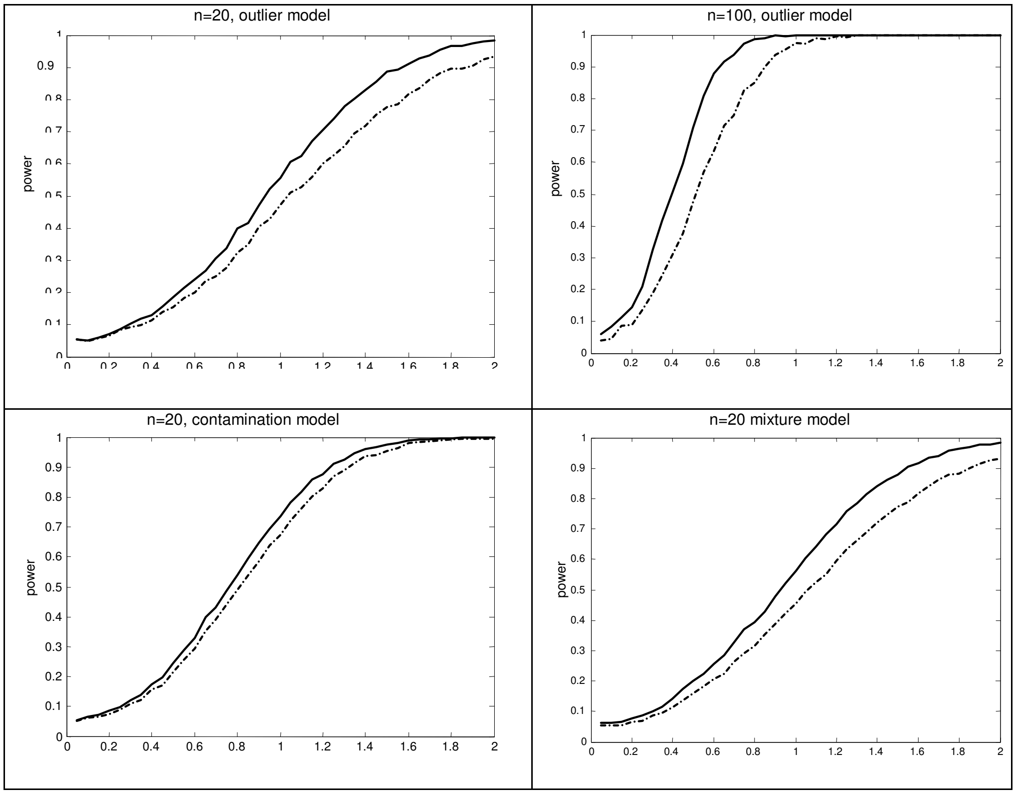

Figure 4.1.2 Power graphs of $\hat{T}^2$ and $\tilde{T}^2$ (GL), data anomalies.

We notice that, even for small n, using $\hat{T}^2$ is advantageous.

4.2 Student's t distributions

4.2.1 The test based on the MML estimators

Before we define the test statistic based on the MMLE we give two lemmas.

Lemma 4.2.1:

(a) The MMLE $\hat{\mu}_x$ is the BAN estimator of μ_x and has asymptotic variance $Var(\hat{\mu}_x) \cong \dfrac{r_x \sigma_1^2}{(r_x+1)m_{11}}$.

(b) The MMLE $\hat{\mu}_u$ is the BAN estimator of μ_u and has asymptotic variance $Var(\hat{\mu}_u) \cong \dfrac{r_u \sigma_1^2}{(r_u+1)m_{21}}$.

Proof:

Note that

$$\frac{\partial \ln L^*}{\partial \mu_x} = \frac{(r_x+1)}{\sigma_1 r_x}\sum_{i=1}^{n}(\alpha_{1i} + \beta_{1i} z_{1(i)}) \text{, i.e.,}$$

$$\frac{\partial \ln L^*}{\partial \mu_x} = \frac{(r_x+1)}{\sigma_1^2 r_x}\sum_{i=1}^{n}\beta_{1i}(x_{(i)} - \mu_x) = \frac{(r_x+1)}{\sigma_1^2 r_x}\left[\sum_{i=1}^{n}\beta_{1i}x_{(i)} - m_{11}\mu_x\right].$$

Now,

$$\frac{\partial \ln L}{\partial \mu_x} \cong \frac{\partial \ln L^*}{\partial \mu_x} = \frac{(r_x+1)m_{11}}{\sigma_1^2 r_x}\left[\hat{\mu}_x - \mu_x\right].$$

The modified maximum likelihood equation is asymptotically equivalent to the maximum likelihood equation and from the above form the result follows.

We can do a similar operation for $\hat{\mu}_u$ and write

$$\frac{\partial \ln L}{\partial \mu_u} \cong \frac{\partial \ln L^*}{\partial \mu_u} = \frac{(r_u+1)m_{21}}{\sigma_1^2 r_u}\left[\hat{\mu}_u - \mu_u\right],$$

which implies, for the same reasons as above, that $\hat{\mu}_u$ is BAN and has asymptotic variance of $Var(\hat{\mu}_u) \cong \dfrac{r_u \sigma_1^2}{(r_u+1)m_{21}}$.

Lemma 4.2.2:

(a) The MMLE $\hat{\mu}_{y/x}$ is the BAN estimator of $\mu_{y/x}$ and has asymptotic variance $Var(\hat{\mu}_{y/x}) \cong \dfrac{r_y \sigma_{2.1}^2}{(r_y+1)m_{12}}$.

(b) The MMLE $\hat{\mu}_{v/u}$ is the BAN estimator of $\mu_{v/u}$ and has asymptotic variance $Var(\hat{\mu}_{v/u}) \cong \dfrac{r_v \sigma_{2.1}^2}{(r_v+1)m_{22}}$.

Proof: Note that

$$\frac{\partial \ln L^*}{\partial \mu_{y/x}} = \frac{(r_y+1)}{\sigma_{2.1} r_y} \sum_{i=1}^{n} \sqrt{c_{1[i]}} (\alpha_{2i} + \beta_{2i} z_{2(i)})$$

Assuming $c_{1i} = 1$ we have $\sum_{i=1}^{n} \alpha_i = 0$. Thus,

$$\frac{\partial \ln L^*}{\partial \mu_{y/x}} = \frac{(r_y+1)}{\sigma_{2.1}^2 r_y} \left[\sum_{i=1}^{n} \beta_i y_{(i)} - m_{12}\mu_{y/x} - \theta \sum_{i=1}^{n} \beta_i x_{(i)} \right]$$

$$= \frac{(r_y+1)m_{12}}{\sigma_{2.1}^2 r_y} \left[\bar{y}_{[.]} - \theta\, \bar{x}_{[.]} - \mu_{y/x} \right].$$

For given θ we have:

$$\frac{\partial \ln L}{\partial \mu_{y/x}} \cong \frac{\partial \ln L^{*}}{\partial \mu_{y/x}} = \frac{(r_y + 1)m_{12}}{\sigma_{2.1}^2 r_y}\left[\hat{\mu}_{y/x} - \mu_{y/x}\right].$$

Therefore, $\hat{\mu}_{y/x}$ is BAN and has asymptotic variance $Var(\hat{\mu}_{y/x}) \cong \frac{r_y \sigma_{2.1}^2}{(r_y + 1)m_{12}}$. Since $\hat{\theta}$ converges to θ as n goes to infinity, the conditionality on θ can be removed.

Similarly,

$$\frac{\partial \ln L}{\partial \mu_{v/u}} \cong \frac{\partial \ln L^{*}}{\partial \mu_{v/u}} = \frac{(r_v + 1)m_{22}}{\sigma_{2.1}^2 r_v}\left[\hat{\mu}_{v/u} - \mu_{v/u}\right]$$

and we see that $\hat{\mu}_{v/u}$ is also BAN with asymptotic variance $Var(\hat{\mu}_{v/u}) \cong \frac{r_v \sigma_{2.1}^2}{(r_v + 1)m_{22}}$.

Now, $Var(\hat{\mu}_x - \hat{\mu}_u) \cong \sigma_1^2\left[\frac{r_x}{(r_x + 1)m_{11}} + \frac{r_u}{(r_u + 1)m_{21}}\right]$ and

$$Var(\hat{\mu}_{y/x} - \hat{\mu}_{v/u}) \cong \sigma_{2.1}^2\left[\frac{r_y}{(r_y + 1)m_{12}} + \frac{r_v}{(r_v + 1)m_{22}}\right].$$

Also, the covariance between $\hat{\mu}_x$ and $\hat{\mu}_{y/x}$ is zero since they are orthogonal components. Similarly, the covariance between $\hat{\mu}_u$ and $\hat{\mu}_{v/u}$ is zero. Thus, asymptotically, the distribution of the vector $\sqrt{n}\left(\hat{\mu}_x - \hat{\mu}_u, \hat{\mu}_{y/x} - \hat{\mu}_{v/u}\right)$ under H_0 is bivariate normal with zero mean vector and variance-covariance matrix

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_{xu}^2 & 0 \\ 0 & \hat{\sigma}_{yv}^2 \end{bmatrix}, \text{ where}$$

$$\hat{\sigma}_{xu}^2 = n\hat{\sigma}_1^2 \left[\frac{r_x}{(r_x + 1)m_{11}} + \frac{r_u}{(r_u + 1)m_{21}} \right] \text{ and}$$

$$\hat{\sigma}_{yv}^2 = n\hat{\sigma}_{2.1}^2 \left[\frac{r_y}{(r_y + 1)m_{12}} + \frac{r_v}{(r_v + 1)m_{22}} \right].$$

Now, to test the null hypothesis $H_0 : \begin{bmatrix} \mu_x - \mu_u \\ \mu_{y/x} - \mu_{v/u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, we define the test statistic:

$$\hat{T}^2 = n\left[\hat{\mu}_x - \hat{\mu}_u \quad \hat{\mu}_{y/x} - \hat{\mu}_{v/u}\right] \hat{\Omega}^{-1} \begin{bmatrix} \hat{\mu}_x - \hat{\mu}_u \\ \hat{\mu}_{y/x} - \hat{\mu}_{v/u} \end{bmatrix}. \qquad (4.2.1)$$

The above test reduces to

$$\hat{T}^2 = \frac{n(\hat{\mu}_x - \hat{\mu}_u)^2}{\hat{\sigma}_{xu}^2} + \frac{n(\hat{\mu}_{y/x} - \hat{\mu}_{v/u})^2}{\hat{\sigma}_{yv}^2}.$$

Since $\hat{\sigma}_1$ converges to σ_1 and $\hat{\sigma}_{2.1}$ converges to $\sigma_{2.1}$ as n goes to infinity, the null distribution of $\hat{T}^2$ is asymptotically chi-square with 2 degrees of freedom. Under H_1, asymptotically, the distribution of $\hat{T}^2$ is non-central chi-square with 2 degrees of freedom and non-centrality parameter

$$\tau^2 = \frac{n(\mu_x - \mu_u)^2}{\sigma_{xu1}^2} + \frac{n(\mu_{y/x} - \mu_{v/u})^2}{\sigma_{yv1}^2}, \text{ where}$$

$$\sigma_{xu1}^2 = n\sigma_1^2 \left[\frac{r_x}{(r_x+1)m_{11}} + \frac{r_u}{(r_u+1)m_{21}} \right] \text{ and}$$

$$\sigma_{yv1}^2 = n\sigma_{2.1}^2 \left[\frac{r_y}{(r_y+1)m_{12}} + \frac{r_v}{(r_v+1)m_{22}} \right].$$

4.2.2 The test based on the LS estimators

Note that

$$Var(\tilde{\mu}_x - \tilde{\mu}_u) = Var(\bar{x} - \bar{u}) = Var(\bar{x}) + Var(\bar{u}) = \frac{\sigma_1^2}{n} \left[\frac{r_x}{r_x - 2} + \frac{r_u}{r_u - 2} \right]$$

and

$$\frac{r_y \sigma_{2.1}^2}{r_y - 2} = Var\left(Y_i - \theta\ X_i\right)$$

or $\dfrac{n r_y \sigma_{2.1}^2}{r_y - 2} = \sum_{i=1}^{n} Var\left(Y_i - \theta\ X_i)\right).$

This implies that

$$\frac{r_y \sigma_{2.1}^2}{n(r_y - 2)} = Var\left(\bar{y}_{(.)} - \theta\ \bar{x}_{(.)})\right).$$

Since $\hat{\theta}$ converges to θ as n goes to infinity we have, asymptotically,

$$Var(\tilde{\mu}_{y/x}) = \frac{r_y \sigma_{2.1}^2}{n\ (r_y - 2)}.$$

Also, $Var(\tilde{\mu}_{v/u}) = \dfrac{r_v \sigma_{2.1}^2}{n\ (r_v - 2)}$.

And thus,

$$Var(\tilde{\mu}_{y/x} - \tilde{\mu}_{v/u}) = \frac{\sigma_{2.1}^2}{n}\left[\frac{r_y}{(r_y - 2)} + \frac{r_v}{(r_v - 2)}\right].$$

Thus, for testing H_0 the test based on the LSE is

$$\tilde{T}^2 = n\begin{bmatrix}\tilde{\mu}_x - \tilde{\mu}_u & \tilde{\mu}_{y/x} - \tilde{\mu}_{v/u}\end{bmatrix} \tilde{\Omega}^{-1}\begin{bmatrix}\tilde{\mu}_x - \tilde{\mu}_u \\ \tilde{\mu}_{y/x} - \tilde{\mu}_{v/u}\end{bmatrix}, \text{ where} \tag{4.2.2}$$

$$\tilde{\Omega} = \begin{bmatrix}\tilde{\sigma}_{xu}^2 & 0 \\ 0 & \tilde{\sigma}_{yv}^2\end{bmatrix};$$

$$\tilde{\sigma}_{xu}^2 = \tilde{\sigma}_1^2\left[\frac{r_x}{(r_x - 2)} + \frac{r_u}{(r_u - 2)}\right] \text{ and}$$

$$\tilde{\sigma}_{yv}^2 = \tilde{\sigma}_{2.1}^2\left[\frac{r_y}{\bar{c}_1(r_y - 2)} + \frac{r_v}{\bar{c}_2(r_v - 2)}\right].$$

The distribution of the test statistic $\tilde{T}^2$ is also asymptotically chi-square with 2 degrees of freedom. Note that $\tilde{T}^2$ can be simplified to give

$$\tilde{T}^2 = \frac{n(\tilde{\mu}_x - \tilde{\mu}_u)^2}{\tilde{\sigma}_{xu}^2} + \frac{n(\tilde{\mu}_{y/x} - \tilde{\mu}_{v/u})^2}{\tilde{\sigma}_{yv}^2}.$$

Under H_1, the distribution of $\tilde{T}^2$ is asymptotically noncentral chi-square with noncentrality parameter

$$\lambda^2 = \frac{n(\mu_x - \mu_u)^2}{\sigma_{xu2}^2} + \frac{n(\mu_{y/x} - \mu_{v/u})^2}{\sigma_{yv2}^2}, \text{ where}$$

$$\sigma_{xu2}^2 = \sigma_1^2\left[\frac{r_x}{(r_x - 2)} + \frac{r_u}{(r_u - 2)}\right] \text{ and}$$

$$\sigma_{yv2}^2 = \sigma_{2.1}^2 \left[\frac{r_y}{\bar{c}_1 (r_y - 2)} + \frac{r_v}{\bar{c}_2 (r_v - 2)} \right].$$

Here also for large n, the chi-square distribution gives a fairly accurate approximation to the percentage points of $\hat{T}^2$ and $\tilde{T}^2$. For small sample sizes we use simulations to find the percentage points. Since $\tau^2 > \lambda^2$, the $\hat{T}^2$ test is asymptotically more powerful than the $\tilde{T}^2$ test.

4.2.3 Comparing $\hat{T}^2$ to $\tilde{T}^2$ (Student's t)

We again perform simulations for different values of n. The graphs are given for $r_x = r_y = r_u = r_v = 4$. Similar graphs can be obtained for different degrees of freedom. Here also we have used the simulated percentage points in each case for the probability of type I error (α) to be 0.05 for both tests. The graphs given below are those of the power for testing

$$H_0 : \begin{bmatrix} \mu_x - \mu_u \\ \mu_{y/x} - \mu_{v/u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ vs}$$

$$H_1 : \begin{bmatrix} \mu_x - \mu_u \\ \mu_{y/x} - \mu_{v/u} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \end{bmatrix} \text{ (as a particular case).}$$

The dotted line represents the power values of the $\tilde{T}^2$ test and the solid line represents the power values of the $\hat{T}^2$ test. Note how the solid line stays above the dotted line even for a sample size of 20 which means that the test based on the MMLE is more powerful than the test based on the LSE.

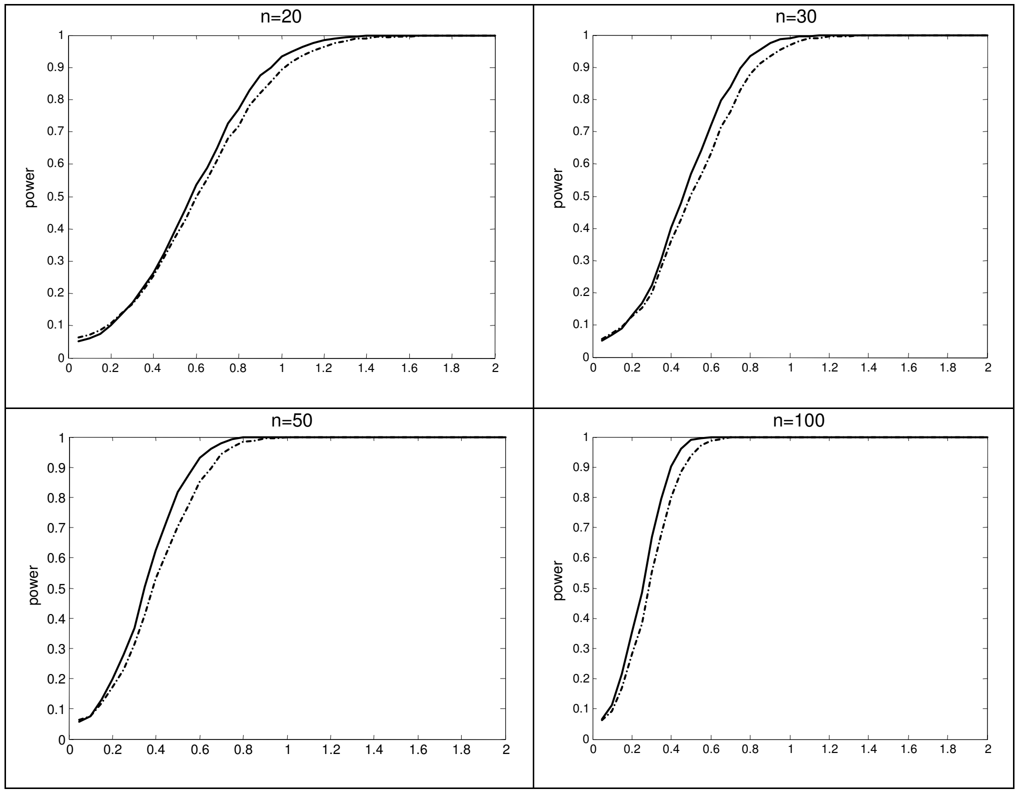

Figure 4.2.1 Power graphs of $\hat{T}^2$ and $\tilde{T}^2$ (Student's t), no data anomalies.

We have also compared the powers of the two test statistics under deviations from the true model. This is shown in Figure 4.2.2. In this case, notice how the power of the test statistic based on the MMLE is much higher than that of the other test. Notice also the big difference between the solid line and the dotted line.

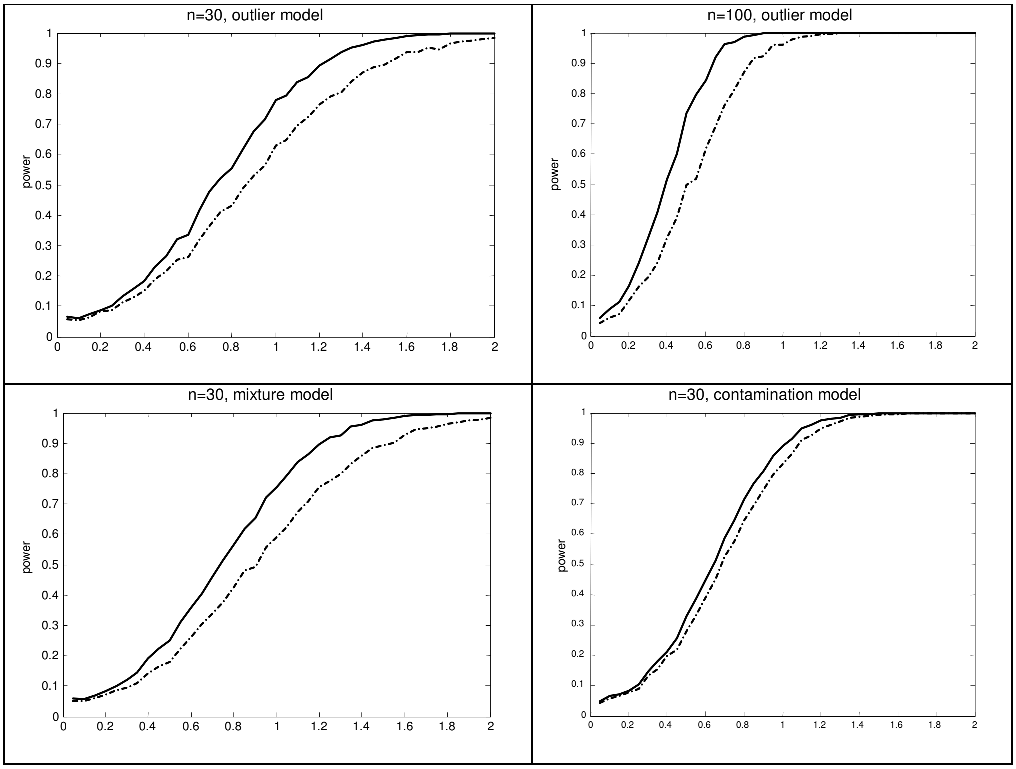

Figure 4.2.2 Power graphs of $\hat{T}^2$ and $\tilde{T}^2$ (Student's t), data anomalies.

Remark: Using the Hotelling type T^2 test based on the MMLE is clearly advantageous.

4.3 Comparing $\hat{T}^2$ to T_D^2

To test that two multivariate distributions are identical, Tiku and Singh (1982) introduced the statistic T_D^2 based on censored normal samples. T_D^2 can in particular be used to test that the population mean vectors are the same. We will use simulations to compare the power of the T_D^2 test to the power of the

$\hat{T}^2$ test for testing the equality of the mean vectors of two bivariate populations.

To test the hypothesis

$$H_0 : \begin{bmatrix} \mu_{11} \\ \mu_{12} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \end{bmatrix} \text{ or equivalently } H_0 : \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} \mu_u \\ \mu_v \end{bmatrix},$$

Tiku and Singh (1982) proposed the statistic

$$T_D^2 = \left(\frac{1}{m_1} + \frac{1}{m_2} \right)^{-1} [\hat{\mu}_1 \quad \hat{\mu}_1] \begin{bmatrix} \hat{\sigma}_1^2 & 0 \\ 0 & \hat{\sigma}_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix}, \qquad (4.3.1)$$

where for the case of symmetric censoring $m_1 = m_2 = n - 2r + 2r\beta = m$,

r = [0.1n + 0.5], $\beta = \frac{-f(t)}{q}(t - f(t)/q)$,

$t = F^{-1}(1-q)$, F is the standard normal CDF, q = r/n,

$\hat{\mu}_1 = \hat{\mu}_{11} - \hat{\mu}_{21}$, $\hat{\mu}_2 = \hat{\mu}_{12} - \hat{\mu}_{22}$,

$\hat{\mu}_{11} = \frac{1}{m}\left[\sum_{i=r+1}^{n-r} x_{(i)} + r\beta(x_{(r+1)} + x_{(n-r)}) \right]$,

$\hat{\mu}_{21} = \frac{1}{m}\left[\sum_{i=r+1}^{n-r} u_{(i)} + r\beta(u_{(r+1)} + u_{(n-r)}) \right]$,

$\hat{\mu}_{12} = \frac{1}{m}\left[\sum_{i=r+1}^{n-r} e_{1(i)} + r\beta(e_{1(r+1)} + e_{1(n-r)}) \right]$; $e_{1i} = y_i - \tilde{\theta}\, x_i$, $\tilde{\theta}$ is the LS estimator of θ,

$\hat{\mu}_{22} = \frac{1}{m}\left[\sum_{i=r+1}^{n-r} e_{2(i)} + r\beta(e_{2(r+1)} + e_{2(n-r)}) \right]$; $e_{2i} = v_i - \tilde{\theta}\, u_i$,

$$\hat{\sigma}_1^2 = \frac{\hat{\sigma}_{11}^2 + \hat{\sigma}_{21}^2}{2}, \; \hat{\sigma}_2^2 = \frac{\hat{\sigma}_{12}^2 + \hat{\sigma}_{22}^2}{2},$$

$$\hat{\sigma}_{11} = \frac{B_x + \sqrt{B_x^2 + 4AC_x}}{2\sqrt{A(A-1)}}; \; A = n - 2r,$$

$$B_x = r\alpha(x_{(n-r)} - x_{(r+1)}); \; \alpha = \frac{f(t)}{q} - \beta t,$$

$$C_x = \sum_{i=r+1}^{n-r} x_{(i)}^2 + r\beta(x_{(r+1)}^2 + x_{(n-r)}^2) - m\hat{\mu}_{11}^2$$

$$\hat{\sigma}_{21} = \frac{B_u + \sqrt{B_u^2 + 4AC_u}}{2\sqrt{A(A-1)}};$$

$$B_u = r\alpha(u_{(n-r)} - u_{(r+1)}), \; C_u = \sum_{i=r+1}^{n-r} u_{(i)}^2 + r\beta(u_{(r+1)}^2 + u_{(n-r)}^2) - m\hat{\mu}_{21}^2,$$

$$\hat{\sigma}_{12} = \frac{B_1 + \sqrt{B_1^2 + 4AC_1}}{2\sqrt{A(A-1)}};$$

$$B_1 = r\alpha(e_{1(n-r)} - e_{1(r+1)}), \; C_1 = \sum_{i=r+1}^{n-r} e_{1(i)}^2 + r\beta(e_{1(r+1)}^2 + e_{1(n-r)}^2) - m\hat{\mu}_{12}^2.$$

$$\hat{\sigma}_{22} = \frac{B_2 + \sqrt{B_2^2 + 4AC_2}}{2\sqrt{A(A-1)}};$$

$$B_2 = r\alpha(e_{2(n-r)} - e_{2(r+1)}), \; C_2 = \sum_{i=r+1}^{n-r} e_{2(i)}^2 + r\beta(e_{2(r+1)}^2 + e_{2(n-r)}^2) - m\hat{\mu}_{22}^2.$$

The rationale Tiku and Singh (1982) used is that “non-normality essentially comes from the tails and, once, extreme observations representing the tails are censored, there is no difference between normal and non-normal samples”.

4.3.1 Generalized Logistic

Here we compare the powers of $\hat{T}^2$ (found in equation (4.1.1)) versus T_D^2 for the Generalized Logistic distribution. Using the simulated percentage points in each case, type I error (α) is 0.05 for both tests. The graphs of the power curves for $\hat{T}^2$ and T_D^2 are given below. The dotted line represents the power values of the T_D^2 test and the solid line represents the power values of the $\hat{T}^2$ test. First we consider the case where $b_x = b_u = b_y = b_v = 1$, which gives an advantage to T_D^2 since the distributions are symmetric and close to normal distributions.

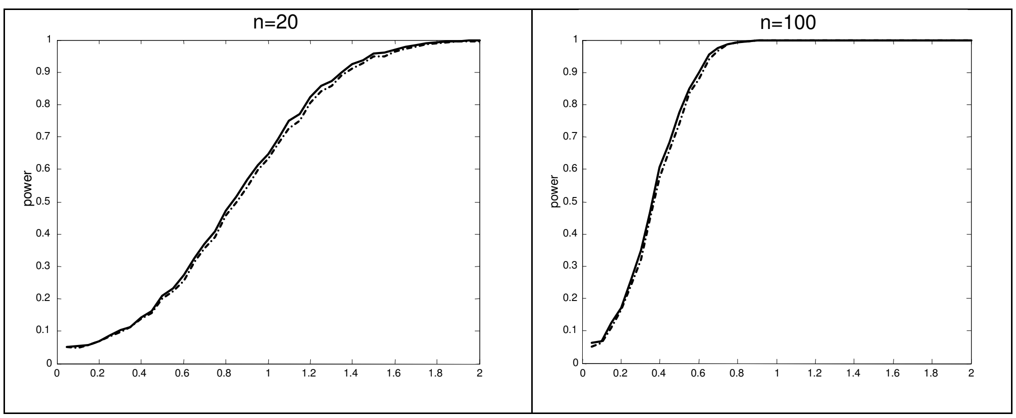

Figure 4.3.1 Power graphs of $\hat{T}^2$ and T_D^2 (GL), $b_x = b_u = b_y = b_v = 1$.

We notice in Figure 4.3.1 that the power curves are close to each other with $\hat{T}^2$ having slightly higher power than T_D^2. However, if we change some of the underlying distributions to be skew, we notice the big difference between the power curves even for small sample sizes. This is shown in Figure 4.3.2.

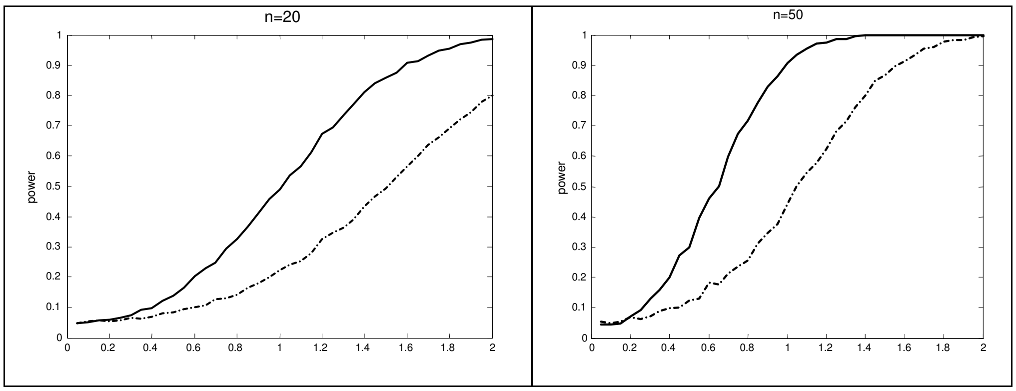

Figure 4.3.2 Power graphs of $\hat{T}^2$ and T_D^2 (GL), $b_x = 0.5, b_u = 1, b_y = 0.5, b_v = 1$.

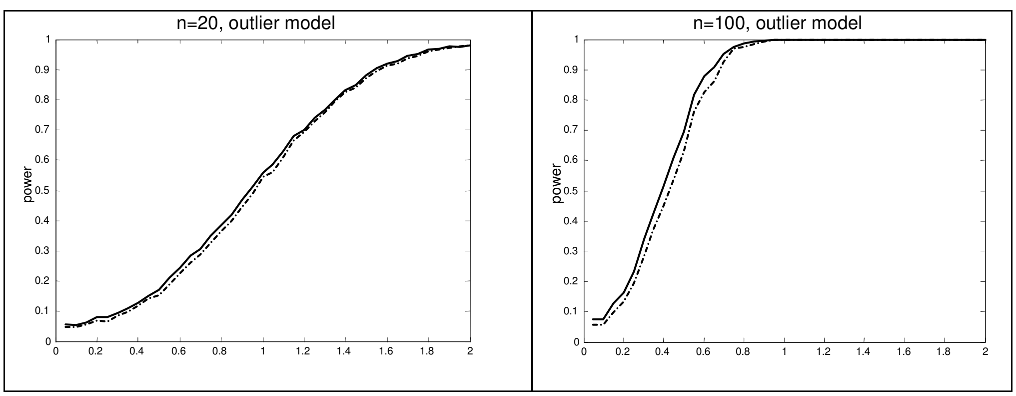

Figure 4.3.3 Power graphs of $\hat{T}^2$ and T_D^2 (GL), $b_x = b_u = b_y = b_v = 1$, outlier model.

Figure 4.3.3 above shows the power graphs of $\hat{T}^2$ and T_D^2 when outliers are present in the data. The outlier model used here is the one mentioned in section 3.1.2 (part c). Notice that $\hat{T}^2$ still has superiority. With the presence of some strong outliers in the data, if $b_x = b_u = b_y = b_v = 1$ and n is small, the power of T_D^2 slightly exceeds that of $\hat{T}^2$. This can be seen in Figure 4.3.4.

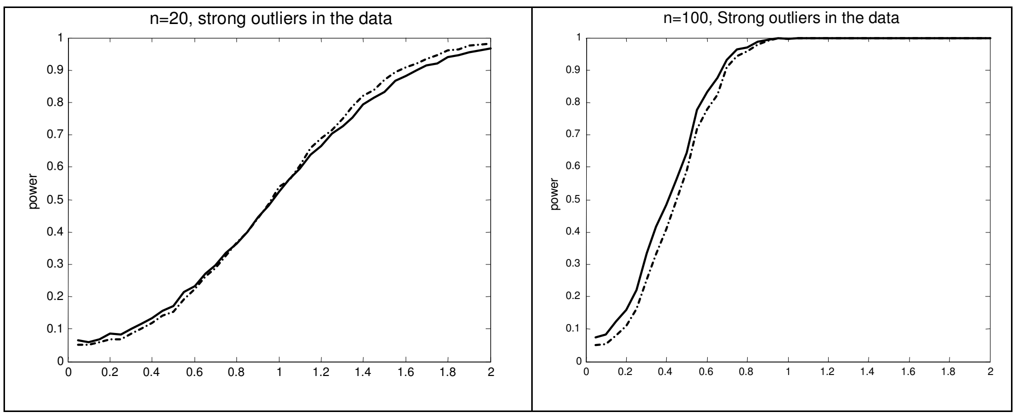

Figure 4.3.4 Power graphs of $\hat{T}^2$ and T_D^2 (GL), $b_x = b_u = b_y = b_v = 1$, with strong outliers in the data.

In the presence of strong outliers, T_D^2 has slightly higher power for the case when $b_x = b_u = b_y = b_v = 1$, however, if we choose some of the underlying generalized logistic distributions to be skew, then $\hat{T}^2$ has much higher power than T_D^2, even in the presence of strong outliers. This is shown in Figure 4.3.5. The outlier model used here is the same as the outlier model mentioned in section 3.1.2 (part c) except that here r of the x_i's come from GL(0, $12\sigma_1$) instead of GL(0, $4\sigma_1$), r of the e_{1i}'s come from GL(0, $12\sigma_{2.1}$)

instead of GL(0, $4\sigma_{2.1}$). Similarly for the u_i's and the e_{2i}'s. Notice the big difference in the power curves in Figure 4.3.5. This clearly shows that using $\hat{T}^2$ is very advantageous.

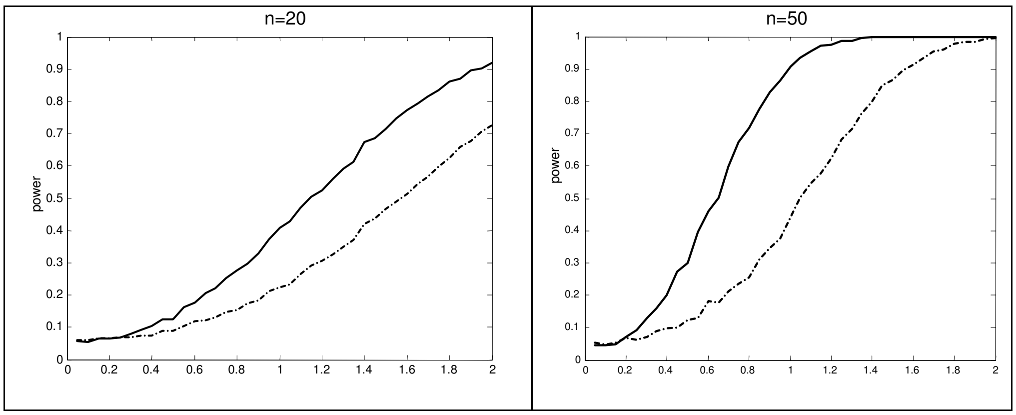

Figure 4.3.5 Power graphs of $\hat{T}^2$ and T_D^2 (GL), $b_x = 0.5, b_u = 1, b_y = 0.5, b_v = 1$, with presence of strong outliers.

4.3.2 Student's t

We will assume the same distribution as in section 4.2. Using the simulated percentage points in each case, type I error (α) is 0.05 for both tests. The graphs of the power curves for $\hat{T}^2$ (given in equation (4.2.1)) and T_D^2 are given below. The dotted line represents the power values of the T_D^2 test and the solid line represents the power values of the $\hat{T}^2$ test. First we assume the degrees of freedom are all equal, i.e. $r_x = r_y = r_u = r_v = 4$.

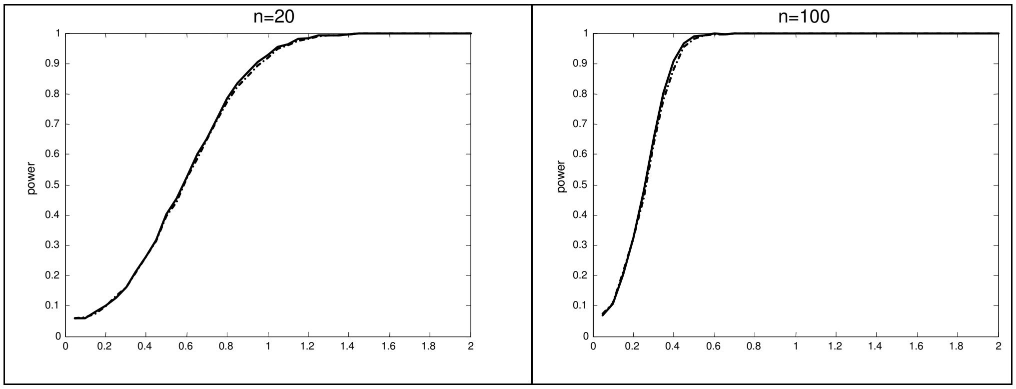

Figure 4.3.6 Power graphs of $\hat{T}^2$ and T_D^2 (Student's t), $r_x = r_y = r_u = r_v = 4$.

Notice that the power curves are almost identical, with the solid line staying above the dotted line slightly. This shows that $\hat{T}^2$ has slightly higher power than T_D^2.

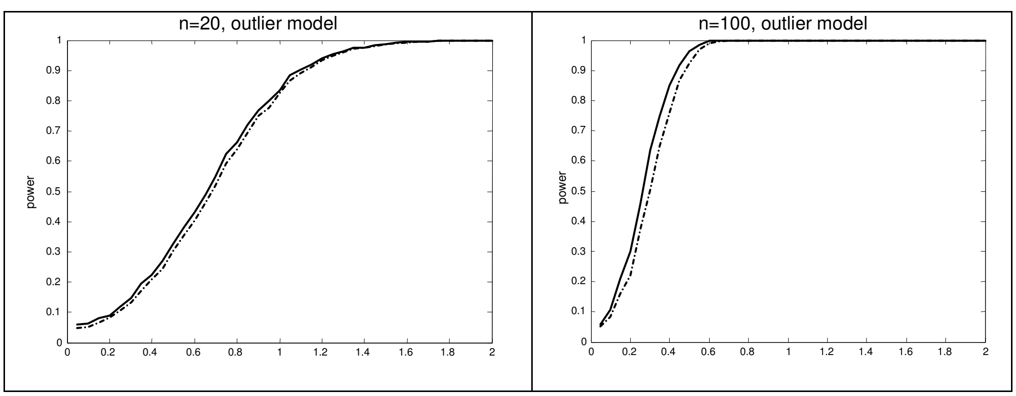

Figure 4.3.7 Power graphs of $\hat{T}^2$ and T_D^2 (Student's t), $r_x = r_y = r_u = r_v = 4$, outlier model.

Figure 4.3.7 shows the power graphs after introducing some outliers in the data. Here we use the outlier model (c) we mentioned in section 3.2.2. We see that the curve corresponding to $\hat{T}^2$ is still slightly above the one that corresponds to T_D^2.

Figure 4.3.8 shows the power graphs after introducing stronger outliers in the data. The outlier model we used is the same as the outlier model mentioned in section 3.2.2 (part c) except that here r of the x_i's come from the distribution given in equation (2.5.1) with σ_1 multiplied by 12 (instead of 4) and r of the e_{1i}'s come from the distribution in equation (2.5.2) with $\sigma_{2.1}$ multiplied by 12. Similarly for the u_i's and the e_{2i}'s.

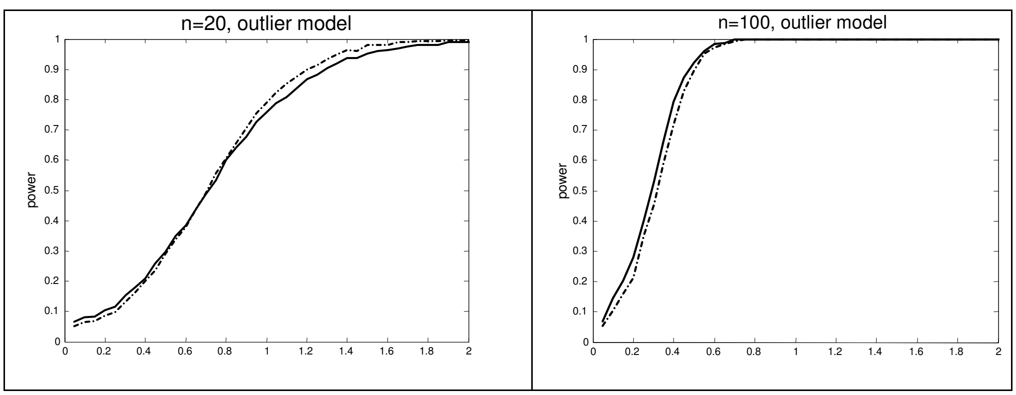

Figure 4.3.8 Power graphs of $\hat{T}^2$ and T_D^2 (Student's t), $r_x = r_y = r_u = r_v = 4$, with the presence of strong outliers.

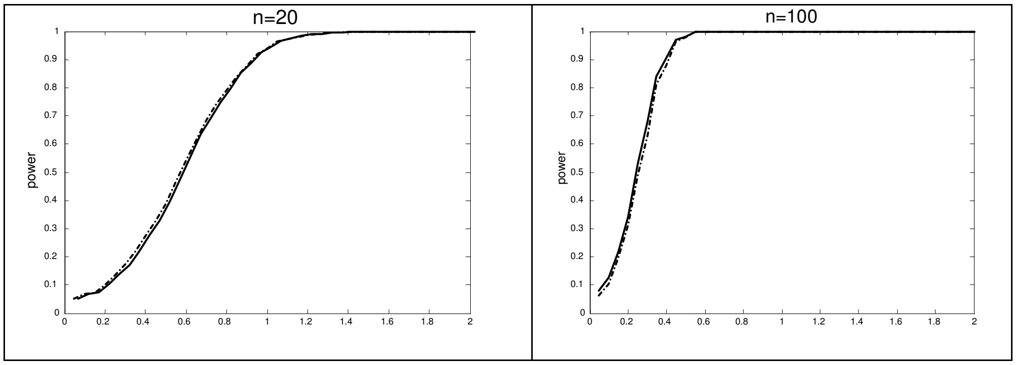

Figure 4.3.9 Power graphs of $\hat{T}^2$ and T_D^2 (Student's t),

$r_x = 4, r_y = 4, r_u = 6, r_v = 6$.

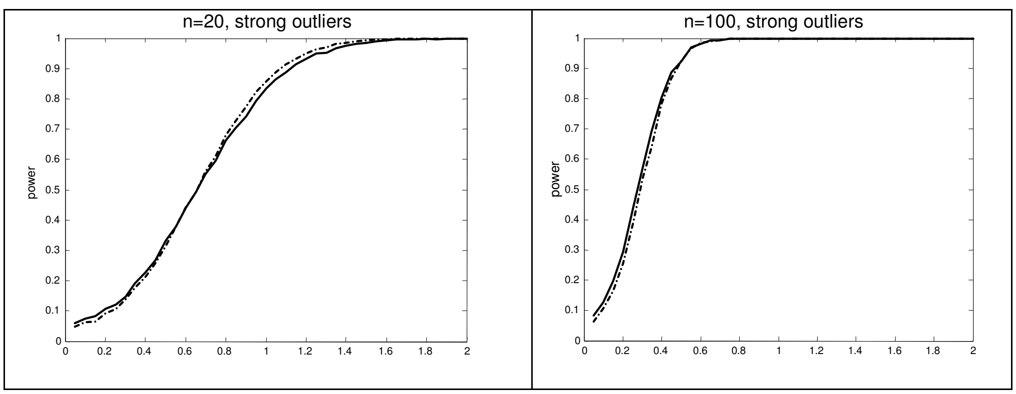

Figure 4.3.10 $r_x = 4, r_y = 4, r_u = 6, r_v = 6$, strong outliers in the data.

Figures 4.3.9 and 4.3.10 show the power graphs for different degrees of freedom. We see that, in presence of strong outliers in the data, T_D^2 has slightly larger power than $\hat{T}^2$ for small values of n. This is because in the calculation of $\hat{T}^2$, outliers are given small weights whereas in the calculation of T_D^2, outliers are given zero weights (censored). However, for large n, $\hat{T}^2$ is always superior over T_D^2 even with the presence of strong outliers. The only time T_D^2 tends to have more power than $\hat{T}^2$ is when the sample size is small, the distribution is symmetric and at the same time there are strong outliers in the data. Even if all three conditions are satisfied, however, the difference between the powers is very small. We have seen that for the generalized logistic, in which case the underlying distributions are skew, $\hat{T}^2$ is much more powerful than T_D^2. Another thing to note is that while T_D^2 may provide a somewhat good way to test hypotheses, the estimators used in the calculation of T_D^2 (based on censored samples) are not as good as the MML estimators we developed in Chapter 2 since the former may have substantial bias. Using $\hat{T}^2$ to test the hypothesis of equal means is, therefore, advantageous in case of symmetric or skew distributions and in presence of outliers as well.

4.4 Mahalanobis Distance

Suppose we have two bivariate populations such that the mean vector for population 1 is $\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$ and the mean vector for population 2 is $\begin{bmatrix} \mu_u \\ \mu_v \end{bmatrix}$. If both populations have a common variance-covariance matrix Ω, the Mahalanobis distance between the two populations is defined as

$$D^2 = \begin{bmatrix} \mu_x - \mu_u & \mu_y - \mu_v \end{bmatrix} \Omega^{-1} \begin{bmatrix} \mu_x - \mu_u \\ \mu_y - \mu_v \end{bmatrix}, \text{ where}$$

$\Omega = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ is the common variance-covariance matrix of the two populations.

Thus, D^2 can be written as

$$D^2 = \begin{bmatrix} \mu_x - \mu_u & \mu_y - \mu_v \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \mu_x - \mu_u \\ \mu_y - \mu_v \end{bmatrix}. \qquad (4.4.1)$$

Now,

$$\Omega^{-1} = \begin{bmatrix} \dfrac{1}{\sigma_1^2 (1-\rho^2)} & \dfrac{-\rho}{\sigma_1 \sigma_2 (1-\rho^2)} \\ \\ \dfrac{-\rho}{\sigma_1 \sigma_2 (1-\rho^2)} & \dfrac{1}{\sigma_1^2 (1-\rho^2)} \end{bmatrix}.$$

Therefore, D^2 can be written as

$$D^2 = \frac{(\mu_x - \mu_u)^2}{\sigma_1^2 (1-\rho^2)} - \frac{2\rho(\mu_x - \mu_u)(\mu_y - \mu_v)}{\sigma_1 \sigma_2 (1-\rho^2)} + \frac{(\mu_y - \mu_v)^2}{\sigma_2^2 (1-\rho^2)}. \qquad (4.4.2)$$

Consider

$$D^{*2} = \begin{bmatrix} \mu_x - \mu_u & \mu_{y/x} - \mu_{v/u} \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_{2.1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \mu_x - \mu_u \\ \mu_{y/x} - \mu_{v/u} \end{bmatrix}.$$

We will show that D^{*2} is equal to D^2.

D^{*2} can be written as

$$D^{*2} = \frac{(\mu_x - \mu_u)^2}{\sigma_1^2} + \frac{(\mu_{y/x} - \mu_{v/u})^2}{\sigma_{2.1}^2}. \qquad (4.4.3)$$

Substituting $\mu_{y/x} = \mu_y - \theta\, \mu_x$, and $\mu_{v/u} = \mu_v - \theta\, \mu_u$, equation (4.4.3) becomes

$$D^{*2} = \frac{(\mu_x - \mu_u)^2}{\sigma_1^2} + \frac{(\mu_y - \mu_v - \theta(\mu_x - \mu_u))^2}{\sigma_{2.1}^2}$$

$$= \frac{(\mu_x - \mu_u)^2}{\sigma_1^2} + \frac{\rho^2(\mu_x - \mu_u)^2}{\sigma_1^2(1-\rho^2)} + \frac{(\mu_y - \mu_v)^2}{\sigma_{2.1}^2} - \frac{2\rho(\mu_x - \mu_u)(\mu_y - \mu_v)}{\sigma_1\sigma_2(1-\rho^2)}$$

$$= \frac{(\mu_x - \mu_u)^2}{\sigma_1^2(1-\rho^2)} - \frac{2\rho(\mu_x - \mu_u)(\mu_y - \mu_v)}{\sigma_1\sigma_2(1-\rho^2)} + \frac{(\mu_y - \mu_v)^2}{\sigma_2^2(1-\rho^2)}$$

$$= D^2 .$$

Let

$$\hat{D}^2 = \begin{bmatrix} \hat{\mu}_x - \hat{\mu}_u & \hat{\mu}_{y/x} - \hat{\mu}_{v/u} \end{bmatrix} \begin{bmatrix} \hat{\sigma}_1^2 & 0 \\ 0 & \hat{\sigma}_{2.1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mu}_x - \hat{\mu}_u \\ \hat{\mu}_{y/x} - \hat{\mu}_{v/u} \end{bmatrix} \text{ and}$$

$$\tilde{D}^2 = \begin{bmatrix} \tilde{\mu}_x - \tilde{\mu}_u & \tilde{\mu}_{y/x} - \tilde{\mu}_{v/u} \end{bmatrix} \begin{bmatrix} \tilde{\sigma}_1^2 & 0 \\ 0 & \tilde{\sigma}_{2.1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mu}_x - \tilde{\mu}_u \\ \tilde{\mu}_{y/x} - \tilde{\mu}_{v/u} \end{bmatrix};$$

$\hat{D}^2$ and $\tilde{D}^2$ are two estimators of D^2. The first one is obtained by replacing all the parameters in D^2 by their MMLE, found in sections 2.4.1 (GL) or 2.5.1 (Student's t). The second one is obtained by replacing all the parameters by their LSE, found in sections 2.4.3 (GL) or 2.5.3 (Student's t). Both $\hat{D}^2$ and $\tilde{D}^2$ converge to D^2 as n tends to infinity.

We will show that $\hat{D}^2$ is more effective in estimating a given distance. To do this we simulate the means and variances of $\hat{D}^2$ and $\tilde{D}^2$. Without loss of generality, as we compare $\hat{D}^2$ to $\tilde{D}^2$, we will assume that the difference between the two population mean vectors is the unit vector. Thus, for all the tables given below we assume that the true values of the parameters are as follows:

$\sigma_1 = 1.0$, $\sigma_2 = 1.0$, $\mu_x = 2.0$, $\mu_y = 2.0$, $\mu_u = 1.0$ and $\mu_v = 1.0$.

In the tables below we give the results of the simulated means and variances of $\hat{D}^2$ and $\tilde{D}^2$ for three representative values of ρ, namely, ρ = 0, 0.5, and 0.9. Note that given the above assumed values of the parameters, we can write $\mu_{y/x}$, $\mu_{v/u}$ and $\sigma_{2.1}$ as functions of ρ:

$$\mu_{y/x} = \mu_y - \frac{\rho\sigma_2}{\sigma_1}\mu_x = 2 - 2\rho = 2(1-\rho),$$

$$\mu_{v/u} = \mu_v - \frac{\rho\sigma_2}{\sigma_1}\mu_u = 1-\rho, \text{ and}$$

$$\sigma_{2.1} = \sigma_2\sqrt{1-\rho^2} = \sqrt{1-\rho^2}.$$

Using the above assumed values, we can write the true value of Mahalanobis distance as a function of ρ as follows:

$$D^2 = \frac{(\mu_x - \mu_u)^2}{\sigma_1^2} + \frac{(\mu_{y/x} - \mu_{v/u})^2}{\sigma_{2.1}^2}$$

$$= \frac{(2-1)^2}{1^2} + \frac{(2(1-\rho)-(1-\rho))^2}{(1-\rho^2)} = \frac{2}{1+\rho}.$$

Remark: In the tables below, $\hat{D}_1^2 = a\hat{D}^2 - b$ and, similarly, $\tilde{D}_1^2$;

$a = \frac{(2n-5)}{2(n-1)}$ and $b = n/4$.

Note: In Tables 4.4.1-4.4.3 we assume $b_x = b_y = b_u = b_v = 1$.

Table 4.4.1 Simulated means and variance of $\hat{D}_1^2$ and $\tilde{D}_1^2$

(GL); ρ = 0. True value of $D^2 = 2$.

n = 20	**Mean**	**Variance**	**MSE**
$\hat{D}_1^2$	2.022	2.434	2.434
$\tilde{D}_1^2$	2.456	3.586	3.794
n=50			
$\hat{D}_1^2$	2.034	1.109	1.110
$\tilde{D}_1^2$	2.190	1.365	1.402
n = 100			
$\hat{D}_1^2$	2.061	0.554	0.558
$\tilde{D}_1^2$	2.132	0.636	0.654

Table 4.4.2 (GL) ρ = 0.5. True value of $D^2 = 1.33$.

n = 20	**Mean**	**Variance**	**MSE**
$\hat{D}_1^2$	1.485	1.784	1.808
$\tilde{D}_1^2$	1.805	2.568	2.794
n=50			
$\hat{D}_1^2$	1.419	0.755	0.762
$\tilde{D}_1^2$	1.527	0.922	0.961
n = 100			
$\hat{D}_1^2$	1.408	0.373	0.379
$\tilde{D}_1^2$	1.455	0.429	0.444

Table 4.4.3 (GL) ρ = 0.9. True value of $D^2 = 1.02$.

n = 20	**Mean**	**Variance**	**MSE**
$\hat{D}_1^2$	1.258	1.535	1.592
$\tilde{D}_1^2$	1.530	2.180	2.440
n=50			
$\hat{D}_1^2$	1.160	0.607	0.627
$\tilde{D}_1^2$	1.249	0.737	0.789
n = 100			
$\hat{D}_1^2$	1.127	0.299	0.310
$\tilde{D}_1^2$	1.164	0.343	0.364

Note: In Tables 4.4.4-4.4.7 we assume that $r_x = r_y = r_u = r_v = 4$.

Table 4.4.4 Simulated means and variance of $\hat{D}_1^2$ and $\tilde{D}_1^2$

(Student's t); ρ = 0. True value of $D^2 = 2$.

n = 20	**Mean**	**Variance**	**MSE**
$\hat{D}_1^2$	1.495	0.988	1.244
$\tilde{D}_1^2$	2.409	2.681	2.848
n=50			
$\hat{D}_1^2$	1.735	0.473	0.543
$\tilde{D}_1^2$	2.215	0.976	1.022
n = 100			
$\hat{D}_1^2$	1.816	0.236	0.269
$\tilde{D}_1^2$	2.111	0.419	0.431

Table 4.4.5 (Student's t) ρ = 0.5. True value of D^2 = 1.33.

n = 20	**Mean**	**Variance**	**MSE**
$\hat{D}_1^2$	1.019	0.668	0.765
$\tilde{D}_1^2$	1.683	1.788	1.913
n=50			
$\hat{D}_1^2$	1.160	0.281	0.310
$\tilde{D}_1^2$	1.483	0.590	0.613
n = 100			
$\hat{D}_1^2$	1.234	0.165	0.174
$\tilde{D}_1^2$	1.432	0.304	0.314

Table 4.4.6 (Student's t) ρ = 0.9. True value of D^2 = 1.02.

n = 20	**Mean**	**Variance**	**MSE**
$\hat{D}_1^2$	0.823	0.512	0.551
$\tilde{D}_1^2$	1.375	1.331	1.457
n=50			
$\hat{D}_1^2$	0.927	0.220	0.229
$\tilde{D}_1^2$	1.197	0.457	0.488
n = 100			
$\hat{D}_1^2$	0.985	0.129	0.130
$\tilde{D}_1^2$	1.135	0.229	0.242

Remark: The values given in Tables 4.2.1-4.4.6 are corrected for the fact that under bivariate normality, $a\tilde{D}^2$ is distributed as noncentral F with degrees of freedom (2, 2n-3) and noncentrality parameter D^2 given in equation (4.4.2).

Notice that in all the above tables, $\hat{D}^2$ has smaller bias and smaller variance than $\tilde{D}^2$ even for small n. Thus, $\hat{D}^2$ is more precise than $\tilde{D}^2$ in estimating

Mahalanobis distance D^2. For bivariate normal populations, $\hat{D}^2$ reduces to $\tilde{D}^2$.

4.5 Testing the Correlation Coefficient

It is of great interest to test the null hypothesis $\rho = 0$. Since the MMLE are asymptotically equivalent to the MLE, in order to test the hypothesis H_0: $\rho = 0$ versus H_1: $\rho \neq 0$, we define the test statistic

$$t^2 = \hat{\rho}^2 / V , \tag{4.5.1}$$

where V is the variance of $\hat{\rho}$ under H_0 and is the last element in the matrix $\hat{I}^{-1}$; I(μ_x, μ_u, σ_1, μ_y, μ_v, σ_2, ρ) being the Fisher information matrix obtained in section 2.4.2 (for the GL distribution) and in section 2.5.2 (for the Student's t distribution). $\hat{I}$ is obtained by replacing σ_1, σ_2 and ρ in I by $\hat{\sigma}_1, \hat{\sigma}_2$ and 0 respectively. The asymptotic null distribution of t^2 is chi-square with 1 degree of freedom. For small n, the null distribution of t^2 is referred to chi-square with 1 degree of freedom.

Based on the LS estimators we define the following test statistic to test H_0

$$t_1^2 = \tilde{\rho}^2 / V_1 . \tag{4.5.2}$$

Since it is difficult to find the variance of $\tilde{\rho}$, we take V_1 to be the simulated variance of $\tilde{\rho}$ for $\rho = 0$. Again, we refer the null distribution of t_1^2 to chi-square distribution with 1 degree of freedom.

Remark 1: In order to test the one sided alternative $H_1:\ \rho > 0$ or $\rho < 0$, we use the statistics $t = \hat{\rho}/\sqrt{V}$ and $t_1 = \tilde{\rho}/\sqrt{V_1}$ based on the MML and the LS estimators, respectively. The null distribution of both test statistics is approximately standard normal.

We compare the powers of the above tests, based on t^2 and t_1^2. We consider Generalized Logistic and Student's t distributions. We graph the power curves for each test. Before we give the power graphs, we list in the table below the simulated 95% points of t^2 and t_1^2 for different values of n. It is interesting to see that they are close to the asymptotic value which is the upper 5% point of the chi-square distribution with one degree of freedom, namely $\chi^2_{0.05}(1) = 3.841$.

Table 4.5.1 Simulated 95% points of t^2 and t_1^2.

Distribution		**n = 60**	**n = 100**	**n = 200**
Generalized Logistic	t^2	4.00	3.31	4.08
	t_1^2	4.05	3.20	3.65
Student's t	t^2	3.90	3.88	3.89
	t_1^2	3.92	3.76	3.77

4.5.1 Generalized Logistic

The graphs of the power curves of t^2 and t_1^2 tests are given below. The dotted line represents the power values of the t_1^2 test and the solid line represents the power values of the t^2 test. We carry out simulations for different values of n assuming that $b_x = b_u = b_y = b_v = 1$, which gives an advantage to t_1^2 since the distributions are symmetric. Note that by using the simulated percentage points in each case, we made sure that type I error (α) is 0.05 for both tests.

The graphs given below represent the power for testing

$$H_0: \rho = 0 \text{ vs. } H_1: \rho \neq 0.$$

Each graph shows a plot of the power against different values of ρ. Note how the solid line stays above the dotted line even for the sample size n = 20 which implies that the test based on the MMLE is more powerful than that based on the LSE.

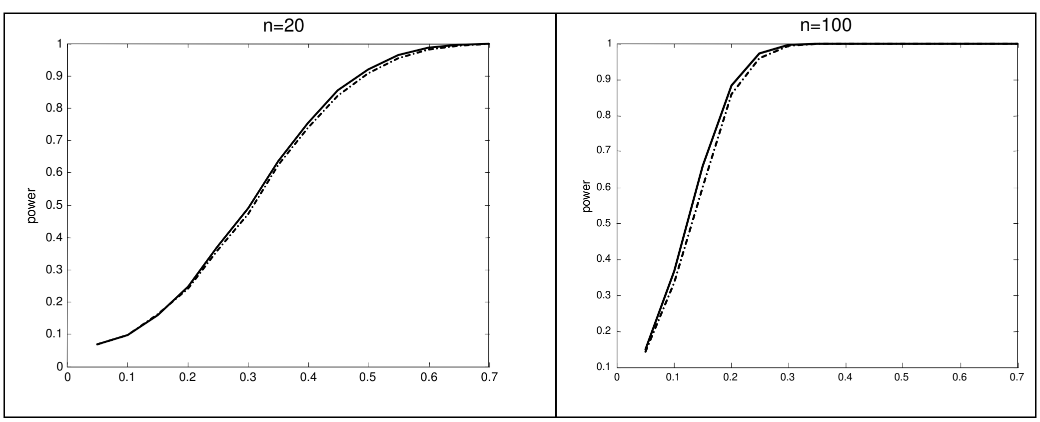

Figure 4.5.1 Power graphs of t^2 and t_1^2 (GL), $b_x = b_u = b_y = b_v = 1$.

In the figures below we show the power graphs of t^2 and t_1^2 test for some deviations from the assumed model. The outlier and contamination models used here are given in section 3.1.2 (part c). We show a few representative graphs. The graphs we show are for $b_x = b_u = b_y = b_v = 1$. Similar graphs can be obtained for different values of n and for different values of b_x, b_u, b_y and b_v. Notice that the solid line is much higher than the dotted line when there are deviations from the assumed distribution especially in case of the outlier model as can be seen in Figure 4.5.2. This illustrates the robustness feature of the t^2 test as compared to the t_1^2 test.

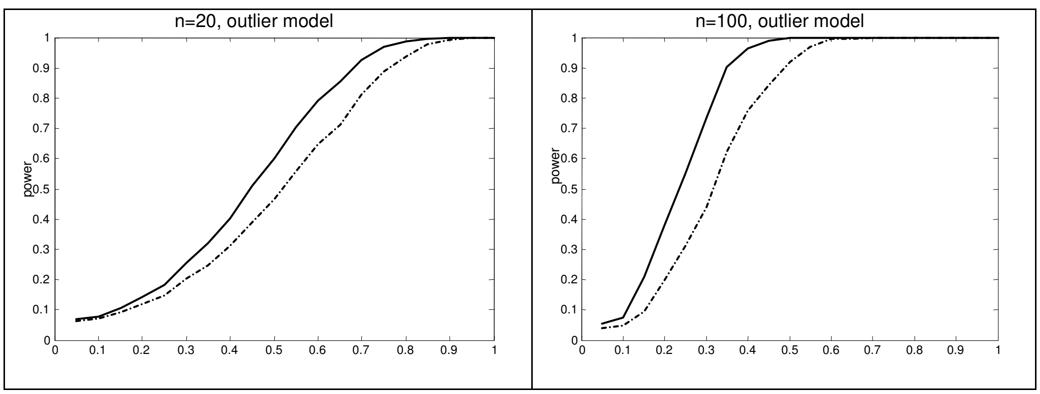

Figure 4.5.2 Power graphs of t^2 and t_1^2 (GL), outlier model.

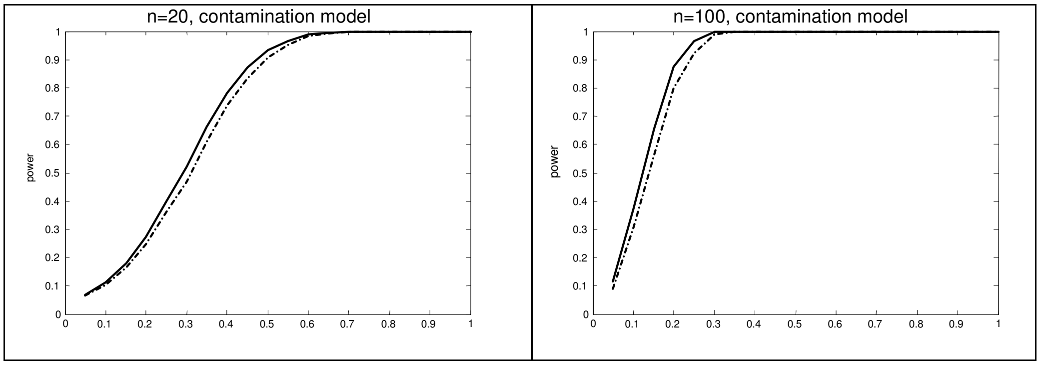

Figure 4.5.3 Power graphs of t^2 and t_1^2 (GL), contamination model.

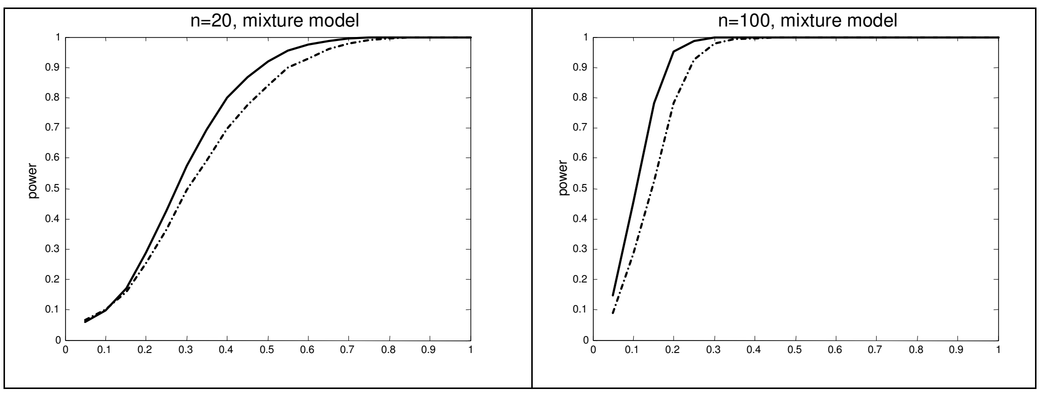

Figure 4.5.4 Power graphs of t^2 and t_1^2 (GL), mixture model.

4.5.2 Student's t

We simulate the power values of the t^2 and t_1^2 tests for $r_x = r_u = r_y = r_v = 4$. The dotted line represents the power of the t_1^2 test and the solid line represents the power of the t^2 test. We carry out simulations for different values of n; the probability of type I error (α) is 0.05 for both tests. Similar graphs can be obtained for different degrees of freedom.

As we notice in Figure 4.5.5, the dotted line is slightly above the solid line only for small sample sizes. For large sample sizes, the power corresponding to the t^2 test is always substantially higher.

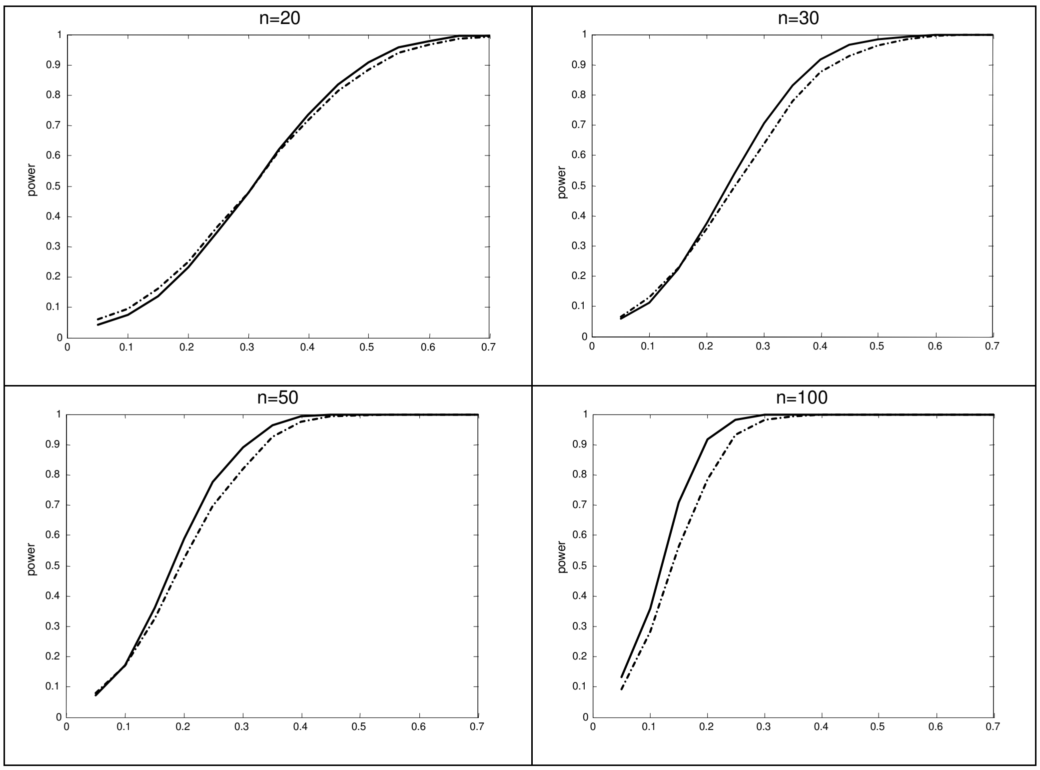

Figure 4.5.5 Power graphs of t^2 and t_1^2 (Student's t).

In the case of deviations from the assumed distribution, or in the presence of outliers in the data, the t^2 test has much higher power as shown in Figure 4.5.6. The outlier, contamination, and mixture models here are the same as in section 3.2.2 (part c).

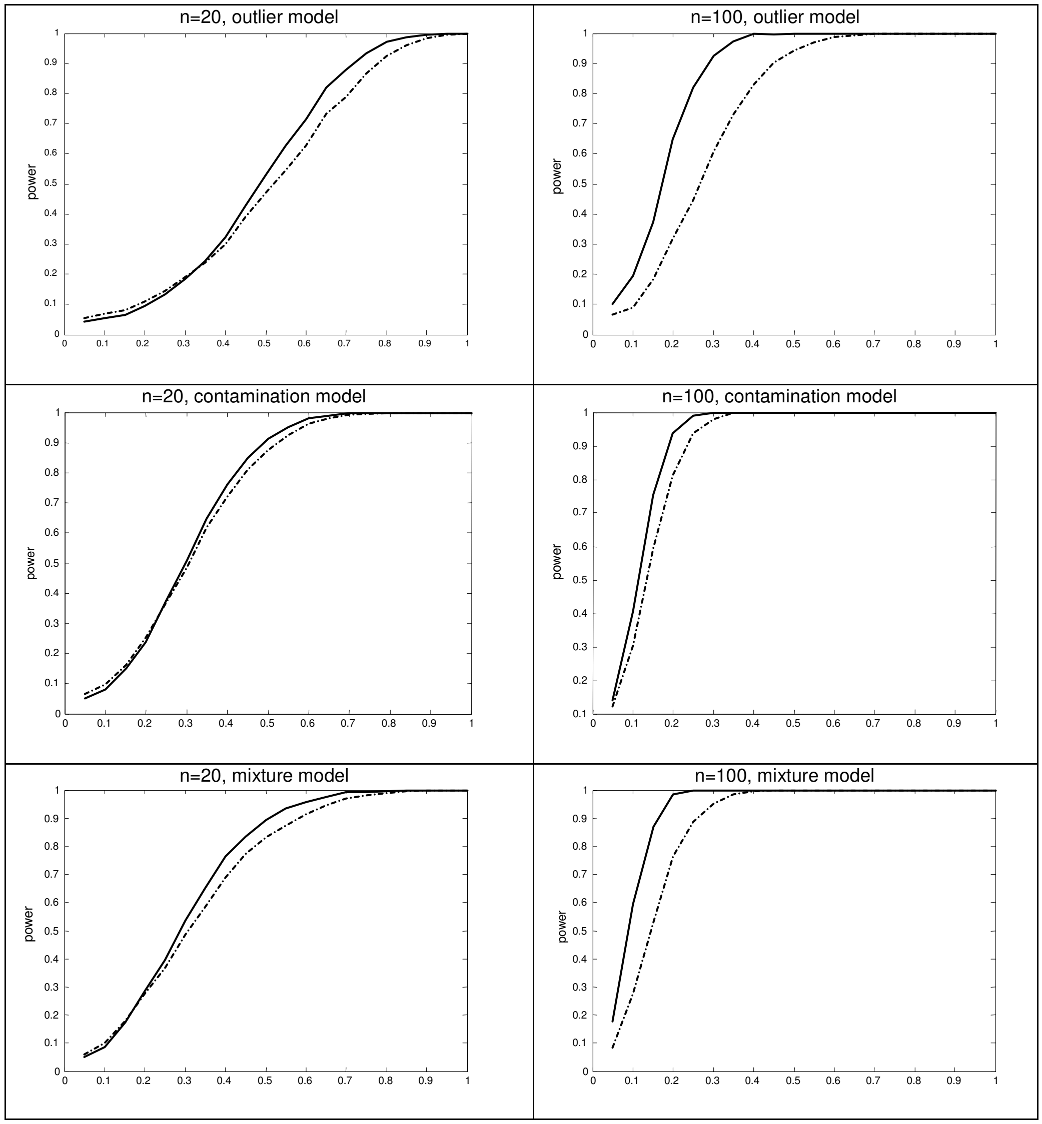

Figure 4.5.6 Power graphs of t^2 and t_1^2 (Student's t), data anomalies.

4.6 Testing the Equality of Two Correlation Coefficients

Suppose that the covariance matrices of the two populations are not the same. The models in this case are given in equations (2.1.1) and (2.1.2), namely,

Model 1: $E(Y/X=x)=\mu_y+\rho_{yx}\dfrac{\sigma_y}{\sigma_x}(x-\mu_x)=\mu_{y/x}+\theta_{yx}x$.

Model 2: $E(V/U=u)=\mu_v+\rho_{vu}\dfrac{\sigma_v}{\sigma_u}(u-\mu_u)=\mu_{v/u}+\theta_{vu}u$.

We would like to test the hypothesis that the correlation coefficients in the two populations are the same, that is, we want to test

$H_0:\ \rho_{yx}=\rho_{vu}$.

Before we define a test statistic to test H_0, we give the estimators of the parameters since they are different than those given in Chapter 2.

4.6.1 Generalized Logistic

In this case the MML and LS estimators of the parameters are as follows (See also Sazak et al. (2006)):

Population 1:

(A) The Modified Maximum Likelihood Estimators:

1. $\hat{\mu}_x=K_{11}-D_{11}\hat{\sigma}_x$ where

$$K_{11} = \frac{\sum \beta_{1i} x_{(i)}}{m_{11}}, \; D_{11} = \frac{1}{m_{11}} \sum \left(\alpha_{1i} - (b_x + 1)^{-1} \right), \text{ and } m_{11} = \sum \beta_{1i} .$$

2. $\hat{\sigma}_x = \dfrac{-B_{11} + \sqrt{B_{11}^2 + 4nC_{11}}}{2\sqrt{n(n-1)}}$ where

$B_{11} = (b_x + 1) \sum \left[\left(\alpha_{1i} - (b_x + 1)^{-1} \right) \left(x_{(i)} - K_{11} \right) \right]$, and

$C_{11} = (b_x + 1) \sum \beta_{1i} (x_{(i)} - K_{11})^2$.

3. $\hat{\mu}_{y/x} = \overline{y}_{[.]} - \hat{\theta}_{yx} \overline{x}_{[.]} - \hat{\sigma}_{2.11} \; \Delta_1 / m_{12}$ where

$\Delta_1 = \sum \Delta_{1i}$, $\Delta_{1i} = (\alpha_{2i} - (b_y + 1)^{-1})$,

$\overline{y}_{[.]} = \dfrac{1}{m_{12}} \sum \beta_{2i} y_{[i]}$, and $\overline{x}_{[.]} = \dfrac{1}{m_{12}} \sum \beta_{2i} x_{[i]}$; $m_{12} = \sum \beta_{2i}$.

4. $\hat{\theta}_{yx} = K_{12} - D_{12} \hat{\sigma}_{2.11}$ where

$K_{12} = \dfrac{\sum \beta_{2i} (x_{[i]} - \overline{x}_{[.]}) y_{(i)}}{\sum \beta_{2i} (x_{[i]} - \overline{x}_{[.]})^2}$ and $D_{12} = \dfrac{\sum \Delta_{1i} (x_{[i]} - \overline{x}_{[.]})}{\sum \beta_{2i} (x_{[i]} - \overline{x}_{[.]})^2}$.

5. $\hat{\sigma}_{2.11} = \dfrac{-B_{12} + \sqrt{B_{12}^2 + 4nC_{12}}}{2\sqrt{n(n-2)}}$ where

$B_{12} = (b_y + 1) \sum \Delta_{1i} (y_{[i]} - \overline{y}_{[.]} - K_{12} (x_{[i]} - \overline{x}_{[.]}))$,

$C_{12} = (b_y + 1) \sum \beta_{1i} (y_{[i]} - \overline{y}_{[.]} - K_{12} (x_{[i]} - \overline{x}_{[.]}))^2$.

6. $\hat{\mu}_y = \overline{y}_{[.]} - \hat{\theta}_{yx} (\overline{x}_{[.]} - \hat{\mu}_x) - \hat{\sigma}_{2.11} \Delta_1 / m_{12}$.

7. $\hat{\sigma}_y = \sqrt{\hat{\sigma}_{2.11}^2 + \hat{\theta}_{yx}^2 \hat{\sigma}_x^2}$.

8. $\hat{\rho}_{yx} = \hat{\theta}_{yx} \dfrac{\hat{\sigma}_x}{\hat{\sigma}_y}$.

(B) The Least Squares Estimators:

1. $\tilde{\mu}_x = \overline{x} - (\psi(b_x) - \psi(1))\tilde{\sigma}_x$.

2. $\tilde{\sigma}_x = \dfrac{s_x}{\sqrt{\psi'(b_x) + \psi'(1)}}$ where

   $s_x^2 = \sum (x_i - \overline{x})^2 / (n-1)$.

3. $\tilde{\mu}_{y/x} = \overline{y} - \tilde{\theta}_{yx} \overline{x} - (\psi(b_y) - \psi(1))\tilde{\sigma}_{2.11}$.

4. $\tilde{\theta}_{yx} = \dfrac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2}$.

5. $\tilde{\sigma}_{2.11} = \sqrt{\dfrac{\left[\sum \left(y_i - \overline{y} - \tilde{\theta}_{yx}(x_i - \overline{x})\right)^2\right] / (n-2)}{\psi'(b_y) + \psi'(1)}}$.

6. $\tilde{\mu}_y = \overline{y} - \tilde{\theta}_{yx} \tilde{\sigma}_x (\psi(b_x) - \psi(1)) - \tilde{\sigma}_{2.11}(\psi(b_y) - \psi(1))$.

7. $\tilde{\sigma}_y = \sqrt{\tilde{\sigma}_{2.11}^2 + \tilde{\theta}_{yx}^2 \tilde{\sigma}_x^2}$.

8. $\tilde{\rho}_{yx} = \dfrac{\tilde{\theta}_{yx}\tilde{\sigma}_x}{\tilde{\sigma}_y}$.

<u>Population 2:</u>

(A) The Modified Maximum Likelihood Estimators

1. $\hat{\mu}_u = K_{21} - D_{21}\hat{\sigma}_u$ where

$$K_{21} = \frac{\sum \gamma_{1i} u_{(i)}}{m_{21}},\ D_{21} = \frac{1}{m_{21}} \sum \left(\delta_{1i} - (b_u + 1)^{-1}\right);\ m_{21} = \sum \gamma_{1i}\ .$$

2. $\hat{\sigma}_u = \dfrac{-B_{21} + \sqrt{B_{21}^2 + 4nC_{21}}}{2\sqrt{n(n-1)}}$ where

$$B_{21} = (b_u + 1)\sum \left[\left(\delta_{1i} - (b_x + 1)^{-1}\right)\left(u_{(i)} - K_{21}\right)\right],$$

$$C_{21} = (b_u + 1)\sum \gamma_{1i}(u_{(i)} - K_{21})^2\ .$$

3. $\hat{\mu}_{v/u} = \bar{v}_{[.]} - \hat{\theta}_{vu}\bar{u}_{[.]} - \hat{\sigma}_{2.12}\Delta_2 / m_{22}$ where

$$\Delta_2 = \sum \Delta_{2i}\ ,\ \ \Delta_{2i} = (\delta_{2i} - (b_v + 1)^{-1}),$$

$$\bar{v}_{[.]} = \frac{1}{m_{22}}\sum \gamma_{2i} v_{[i]}\ ,\ \bar{u}_{[.]} = \frac{1}{m_{22}}\sum \gamma_{2i} u_{[i]}\ ;\ m_{22} = \sum \gamma_{2i}\ .$$

4. $\hat{\theta}_{vu} = K_{22} - D_{22}\hat{\sigma}_{2.12}$ where

$$K_{22} = \frac{\sum \gamma_{2i}(u_{[i]} - \bar{u}_{[.]})v_{(i)}}{\sum \gamma_{2i}(u_{[i]} - \bar{u}_{[.]})^2}, \text{ and } D_{22} = \frac{\sum \Delta_{2i}(u_{[i]} - \bar{u}_{[.]})}{\sum \gamma_{2i}(u_{[i]} - \bar{u}_{[.]})^2}\ .$$

5. $\hat{\sigma}_{2.12} = \dfrac{-B_{22} + \sqrt{B_{22}^2 + 4nC_{22}}}{2\sqrt{n(n-2)}}$ where

$B_{22} = (b_v + 1)\sum \Delta_{2i}(v_{[i]} - \bar{v}_{[.]} - K_{22}(u_{[i]} - \bar{u}_{[.]}))$,

$C_{22} = (b_v + 1)\sum \delta_{1i}(v_{[i]} - \bar{v}_{[.]} - K_{22}(u_{[i]} - \bar{u}_{[.]}))^2$.

6. $\hat{\mu}_v = \bar{v}_{[.]} - \hat{\theta}_{vu}(\bar{u}_{[.]} - \hat{\mu}_u) - \hat{\sigma}_{2.12}\Delta_2 / m_{22}$.

7. $\hat{\sigma}_v = \sqrt{\hat{\sigma}_{2.12}^2 + \hat{\theta}_{vu}^2 \hat{\sigma}_u^2}$.

8. $\hat{\rho}_{vu} = \hat{\theta}_{vu} \dfrac{\hat{\sigma}_u}{\hat{\sigma}_v}$.

(B) The Least Squares Estimators

1. $\tilde{\mu}_u = \bar{u} - (\psi(b_u) - \psi(1))\tilde{\sigma}_u$.

2. $\tilde{\sigma}_u = \dfrac{s_u}{\sqrt{\psi'(b_u) + \psi'(1)}}$ where

$s_u^2 = \sum (u_i - \bar{u})^2 / (n-1)$.

3. $\tilde{\mu}_{v/u} = \bar{v} - \tilde{\theta}_{vu}\bar{u} - (\psi(b_v) - \psi(1))\tilde{\sigma}_{2.12}$.

4. $\tilde{\theta}_{vu} = \dfrac{\sum (u_i - \bar{u})v_i}{\sum (u_i - \bar{u})^2}$.

5. $\tilde{\sigma}_{2.12} = \sqrt{\dfrac{\left[\sum\left(v_i - \bar{v} - \tilde{\theta}_{vu}(u_i - \bar{u})\right)^2\right]/(n-2)}{\psi'(b_v) + \psi'(1)}}$.

6. $\tilde{\mu}_v = \bar{v} - \tilde{\theta}_{vu}\tilde{\sigma}_u(\psi(b_u) - \psi(1)) - \tilde{\sigma}_{2.12}(\psi(b_v) - \psi(1))$.

7. $\tilde{\sigma}_v = \sqrt{\tilde{\sigma}_{2.12}^2 + \tilde{\theta}_{vu}^2\tilde{\sigma}_u^2}$.

8. $\tilde{\rho}_{vu} = \dfrac{\tilde{\theta}_{vu}\tilde{\sigma}_u}{\tilde{\sigma}_v}$.

Note: all the above summations are from 1 to n.

4.6.2 Student's t

Here, the estimators of the parameters are as follows (see also Tiku et al. (2007)):

<u>Population 1</u>:

(A) The Modified Maximum Likelihood Estimators:

1. $\hat{\mu}_x = \dfrac{\sum\beta_{1i}x_{(i)}}{m_{11}}$; $m_{11} = \sum\beta_{1i}$.

2. $\hat{\sigma}_x = \dfrac{B_{11} + \sqrt{B_{11}^{\;2} + 4nC_{11}}}{2\sqrt{n(n-1)}}$ where

$$B_{11} = \frac{(r_x + 1)}{r_x} \sum_{i=1}^{n} \alpha_{1i} x_{(i)} \text{ and } C_{11} = \frac{(r_x + 1)}{r_x} \sum_{i=1}^{n} \beta_{1i} (x_{(i)} - \hat{\mu}_x)^2 .$$

3. $\hat{\mu}_{y/x} = \bar{y}_{[.]} - \hat{\theta}_{yx} \bar{x}_{[.]} - \hat{\sigma}_{2.11} \sum_{i=1}^{n} \alpha_{2i} / m_{12}$ where

$$\bar{y}_{[.]} = \frac{1}{m_{12}} \sum \beta_{2i} y_{[i]} \text{ and } \bar{x}_{[.]} = \frac{1}{m_{12}} \sum \beta_{2i} x_{[i]} ; \; m_{12} = \sum \beta_{2i} .$$

4. $\hat{\theta}_{yx} = K_{12} + D_{12} \hat{\sigma}_{2.11}$ where

$$K_{12} = \frac{\sum \beta_{2i} (x_{[i]} - \bar{x}_{[.]}) y_{(i)}}{\sum \beta_{2i} (x_{[i]} - \bar{x}_{[.]})^2} \text{ and } D_{12} = \frac{\sum \alpha_{2i} (x_{[i]} - \bar{x}_{[.]})}{\sum \beta_{2i} (x_{[i]} - \bar{x}_{[.]})^2} .$$

5. $\hat{\sigma}_{2.11} = \dfrac{B_{12} + \sqrt{B_{12}^2 + 4nC_{12}}}{2\sqrt{n(n-2)}}$ where

$$B_{12} = \frac{(r_y + 1)}{r_y} \sum \alpha_{2i} (y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]})) \text{ and}$$

$$C_{12} = \frac{(r_y + 1)}{r_y} \sum \beta_{2i} (y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]}))^2 .$$

6. $\hat{\mu}_y = \hat{\mu}_{y/x} + \hat{\theta}_{yx} \hat{\mu}_x .$

7. $\hat{\sigma}_y = \sqrt{\hat{\sigma}_{2.11}^2 + \hat{\theta}_{yx}^2 \hat{\sigma}_x^2} .$

8. $\hat{\rho}_{yx} = \hat{\theta}_{yx} \dfrac{\hat{\sigma}_x}{\hat{\sigma}_y} .$

(B) The Least Squares Estimators:

1. $\tilde{\mu}_x = \bar{x}$.

2. $\tilde{\sigma}_x = \sqrt{\dfrac{r_x - 2}{r_x}}\; s_x$ where $s_x^2 = \sum (x_i - \bar{x})^2 / (n-1)$.

3. $\tilde{\mu}_{y/x} = \bar{y} - \tilde{\theta}_{yx}\bar{x}$.

4. $\tilde{\theta}_{yx} = \dfrac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}$.

5. $\tilde{\sigma}_{2.11} = \sqrt{\dfrac{(r_y - 2)}{r_y} \dfrac{\left[\sum \left(y_i - \bar{y} - \tilde{\theta}_{yx}(x_i - \bar{x})\right)^2\right]}{(n-2)}}$.

6. $\tilde{\mu}_y = \tilde{\mu}_{y/x} + \tilde{\theta}_{yx}\bar{x}$.

7. $\tilde{\sigma}_y = \sqrt{\tilde{\sigma}_{2.11}^2 + \tilde{\theta}_{yx}^2 \tilde{\sigma}_x^2}$.

8. $\tilde{\rho}_{yx} = \dfrac{\tilde{\theta}_{yx}\tilde{\sigma}_x}{\tilde{\sigma}_y}$.

Population 2:

(A) The Modified Maximum Likelihood Estimators

1. $\hat{\mu}_u = \dfrac{\sum \gamma_{1i} u_{(i)}}{m_{21}}$ where $m_{21} = \sum \gamma_{1i}$

2. $\hat{\sigma}_u = \dfrac{B_{21} + \sqrt{B_{21}^2 + 4nC_{21}}}{2\sqrt{n(n-1)}}$ where

$$B_{21} = \frac{(r_u + 1)}{r_u} \sum_{i=1}^{n} \delta_{1i} u_{(i)} \text{ and}$$

$$C_{21} = \frac{(r_u + 1)}{r_u} \sum_{i=1}^{n} \gamma_{1i} (u_{(i)} - \hat{\mu}_u)^2 .$$

3. $\hat{\mu}_{v/u} = \bar{v}_{[.]} - \hat{\theta}_{vu} \bar{u}_{[.]} + (\hat{\sigma}_{2.12} / m_{22}) \sum_{i=1}^{n} \delta_{2i}$ where

$$\bar{v}_{[.]} = \frac{1}{m_{22}} \sum_{i=1}^{n} \gamma_i v_{[i]} \text{ , } \bar{u}_{[.]} = \frac{1}{m_{22}} \sum_{i=1}^{n} \gamma_i u_{[i]} \text{ ; } m_{22} = \sum_{i=1}^{n} \gamma_i \text{ .}$$

4. $\hat{\theta}_{vu} = K_{22} + D_{22} \hat{\sigma}_{2.12}$ where

$$K_{22} = \frac{\sum \gamma_{2i} (u_{[i]} - \bar{u}_{[.]}) v_{(i)}}{\sum \gamma_{2i} (u_{[i]} - \bar{u}_{[.]})^2} ,$$

$$D_{22} = \frac{\sum \alpha_{2i} (u_{[i]} - \bar{u}_{[.]})}{\sum \gamma_{2i} (u_{[i]} - \bar{u}_{[.]})^2} .$$

5. $\hat{\sigma}_{2.12} = \dfrac{B_{22} + \sqrt{B_{22}^2 + 4nC_{22}}}{2\sqrt{n(n-2)}}$ where

$$B_{22} = \frac{(r_v + 1)}{r_v} \sum \delta_{2i} (v_{[i]} - \bar{v}_{[.]} - K(u_{[i]} - \bar{u}_{[.]})) ,$$

$$C_{22} = \frac{(r_v + 1)}{r_v} \sum \gamma_{2i} (v_{[i]} - \bar{v}_{[.]} - K(u_{[i]} - \bar{u}_{[.]}))^2 .$$

6. $\hat{\mu}_v = \hat{\mu}_{v/u} + \hat{\theta}_{vu} \hat{\mu}_u$.

7. $\hat{\sigma}_v = \sqrt{\hat{\sigma}^2_{2.12} + \hat{\theta}^2_{vu} \hat{\sigma}^2_u}$.

8. $\hat{\rho}_{vu} = \hat{\theta}_{vu} \dfrac{\hat{\sigma}_u}{\hat{\sigma}_v}$.

(B) The Least Squares Estimators

1. $\tilde{\mu}_u = \bar{u}$.

2. $\tilde{\sigma}_u = \sqrt{\dfrac{r_u - 2}{r_u}} \; s_u$ where $s^2_u = \sum (u_i - \bar{u})^2 / (n-1)$.

3. $\tilde{\mu}_{v/u} = \bar{v} - \tilde{\theta} \bar{u}$.

4. $\tilde{\theta}_{vu} = \dfrac{\sum (u_i - \bar{u}) v_i}{\sum (u_i - \bar{u})^2}$.

5. $\tilde{\sigma}_{2.12} = \sqrt{\dfrac{(r_v - 2)}{r_v} \dfrac{\left[\sum \left(v_i - \bar{v} - \tilde{\theta}_{vu} (u_i - \bar{u})\right)^2\right]}{(n-2)}}$.

6. $\tilde{\mu}_v = \tilde{\mu}_{v/u} + \tilde{\theta}_{vu}\bar{u}$.

7. $\tilde{\sigma}_v = \sqrt{\tilde{\sigma}_{2.12}^2 + \tilde{\theta}_{vu}^2 \tilde{\sigma}_u^2}$.

8. $\tilde{\rho}_{vu} = \dfrac{\tilde{\theta}_{vu}\tilde{\sigma}_u}{\tilde{\sigma}_v}$.

Note: all the above summations are from 1 to n.

Remark 1: In calculating the above MML estimators for both the Generalized logistic and the Student's t, the α's, β's, δ's and γ's are the same as those given in Chapter 2.

Remark 2: The Fisher information matrices when the var-cov matrices of the two populations are not the same are given in Appendix F.

4.6.3 Test Statistics

We will now use the estimators given above to test the hypothesis

$$H_0: \rho_{yx} = \rho_{vu} \quad \text{vs.} \quad H_1: \rho_{yx} \neq \rho_{vu}.$$

In order to test H_0, we note that the MML estimators are asymptotically equivalent to the ML estimators and thus the null distribution of $\hat{\rho}_{yx}$ is asymptotically normal. The same holds for $\hat{\rho}_{vu}$.

Let,

$$Z_1 = 0.5\ln\left[\frac{1+\hat{\rho}_{yx}}{1-\hat{\rho}_{yx}}\right] \text{ and } Z_2 = 0.5\ln\left[\frac{1+\hat{\rho}_{vu}}{1-\hat{\rho}_{vu}}\right].$$

The null distribution of $Z_1 - Z_2$ is asymptotically normal with zero mean and variance which is a constant multiple of $\frac{2}{n-3}$. Thus, we define the statistic:

$$W^2 = \frac{\left\{0.5\ln\left[\frac{1+\hat{\rho}_{yx}}{1-\hat{\rho}_{yx}}\right] - 0.5\ln\left[\frac{1+\hat{\rho}_{vu}}{1-\hat{\rho}_{vu}}\right]\right\}^2}{2/(n-3)}.$$

For large n the asymptotic null distribution of W^2 is chi-square with 1 degree of freedom.

Based on the LS estimators we have the statistic

$$W_1^2 = \frac{\left\{0.5\ln\left[\frac{1+\tilde{\rho}_{yx}}{1-\tilde{\rho}_{yx}}\right] - 0.5\ln\left[\frac{1+\tilde{\rho}_{vu}}{1-\tilde{\rho}_{vu}}\right]\right\}^2}{2/(n-3)}.$$

The null distribution of W_1^2 is also a constant multiple of chi-square with one degree of freedom.

Remark: For testing a one sided alternative H_1: $\rho_{yx} > \rho_{vu}$ or $\rho_{yx} < \rho_{vu}$, we use the test statistics

$$W = \frac{\left\{0.5\ln\left[\frac{1+\hat{\rho}_{yx}}{1-\hat{\rho}_{yx}}\right] - 0.5\ln\left[\frac{1+\hat{\rho}_{vu}}{1-\hat{\rho}_{vu}}\right]\right\}}{\sqrt{2/(n-3)}} \text{ and}$$

$$W_1 = \frac{\left\{0.5\ln\left[\frac{1+\tilde{\rho}_{yx}}{1-\tilde{\rho}_{yx}}\right] - 0.5\ln\left[\frac{1+\tilde{\rho}_{vu}}{1-\tilde{\rho}_{vu}}\right]\right\}}{\sqrt{2/(n-3)}}.$$

The asymptotic null distributions of both W and W_1 are constant multiples of a standard normal.

Before we compare the powers of the two tests, we give in the table below the simulated 95% points of W^2 and W_1^2 for different values of n.

Table 4.6.1 Simulated 95% points of W^2 and W_1^2.

Distribution		**n = 60**	**n = 100**	**n = 200**
Generalized Logistic	W^2	3.90	4.10	4.10
	W_1^2	4.40	4.40	4.90
Student's t	W^2	4.40	4.00	3.80
	W_1^2	6.20	6.60	6.70

For W^2, the percentage points are interestingly close to the upper 5% point of chi-square distribution with one degree of freedom, namely, $\chi^2_{0.05}(1)$ = 3.841. It implies that the null distribution of W^2 is closely approximated by chi-square with 1 degree of freedom, but not the distribution of W_1^2. The latter needs adjusting for the variance which is theoretically difficult (Gayen, 1951).

4.6.4 Power Simulations

We now compare the power of the two tests, namely, W^2 and W_1^2.

We simulate the power of W^2 and W_1^2 for the following hypothesis:

H_0: ρ_{yx} - $\rho_{vu} = 0$ vs. H_1: ρ_{yx} - $\rho_{vu} \neq 0$.

We graph the power curves of W^2 and W_1^2 for different values of ρ.

Generalized Logistic

The graphs of the power of W^2 and W_1^2 tests are given below. The dotted line represents the power of W_1^2 and the solid line represents the power of W^2. We did simulations for different values of n assuming that $b_x = b_u = b_y = b_v = 1$. Note that in the cases where the probability of type I error (α) was not 0.05, we used the simulated percentage points making sure that α is 0.05 for both tests. Deviations from the hypothetical critical value occur mostly in the outlier model.

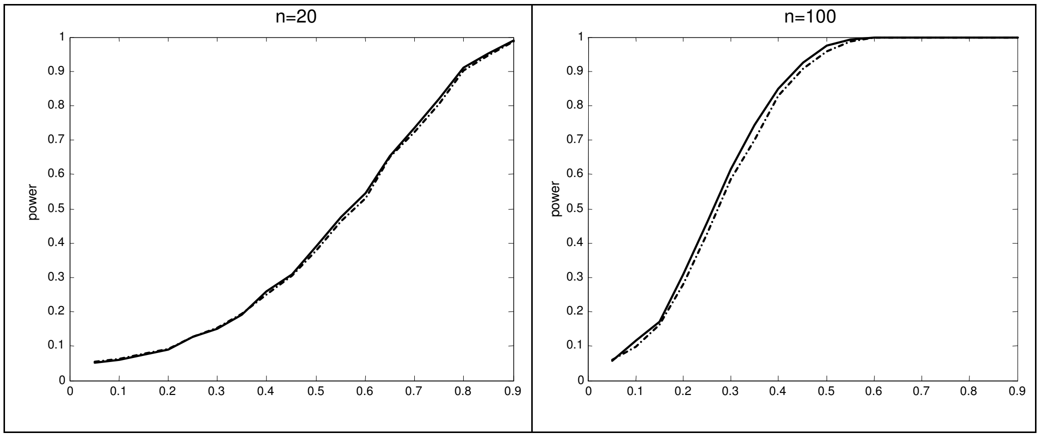

Figure 4.6.1 Power graphs of W^2 and W_1^2 (GL).

In the above graphs we notice that the power of the W^2 test is slightly higher than the power of the W_1^2 test. We also simulated the powers of both tests under deviations from our assumed model. Notice the big difference between the power curves when outliers are present in the data, as shown in Figure 4.6.2. The solid line is much higher than the dotted line. This shows that the test based on W^2 is more robust and powerful as compared to W_1^2. Here the outlier, mixture and contamination models are the same as the ones used in section 3.1.2 (part c).

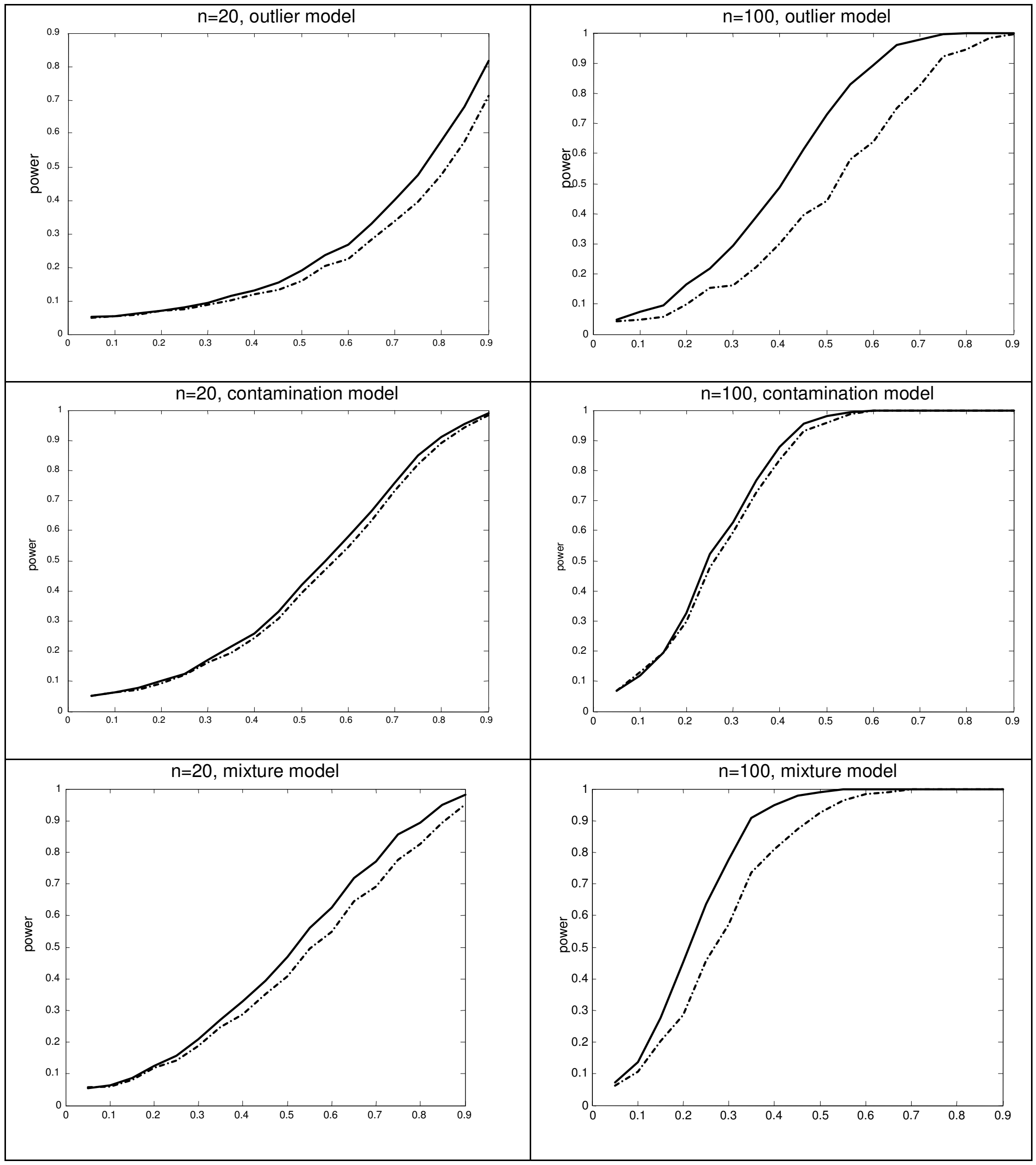

Figure 4.6.2 Power graphs of W^2 and W_1^2 (GL), data anomalies.

Student's t

We give the power curves of the W^2 and W_1^2 tests when $r_x = r_u = r_y = r_v = 4$, type I error being 0.05.

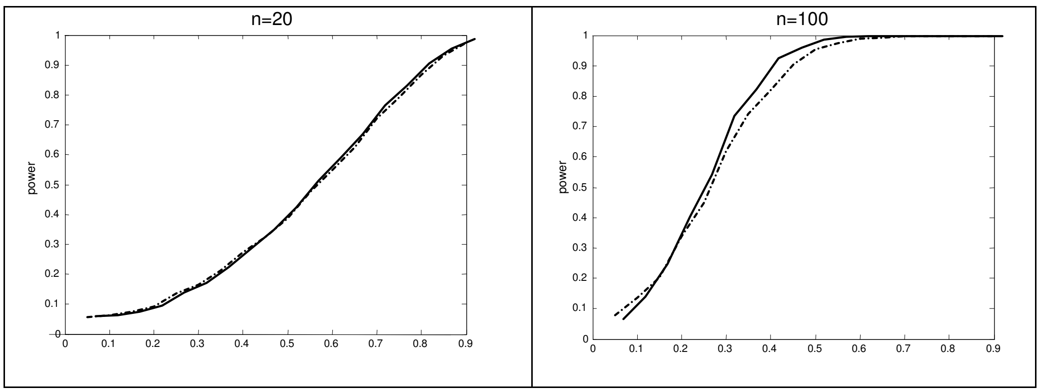

Figure 4.6.3 Power graphs of W^2 and W_1^2 (Student's t).

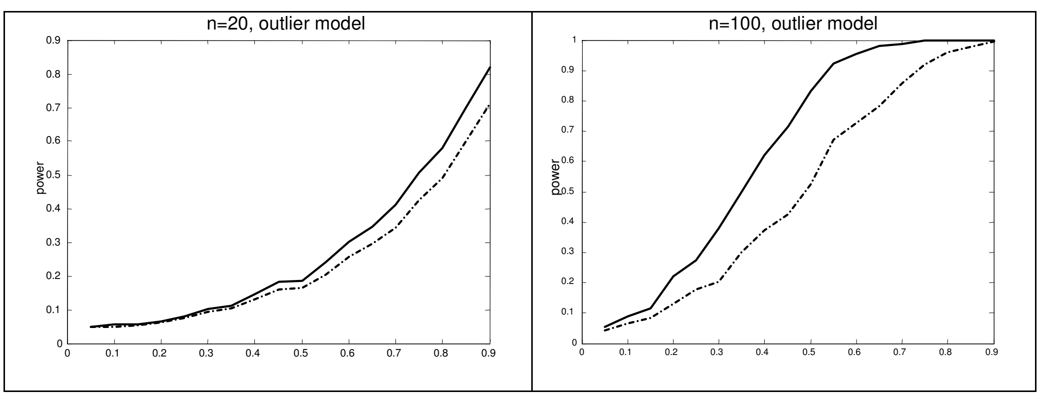

Figure 4.6.4 Power graphs of W^2 and W_1^2 (Student's t), outlier model.

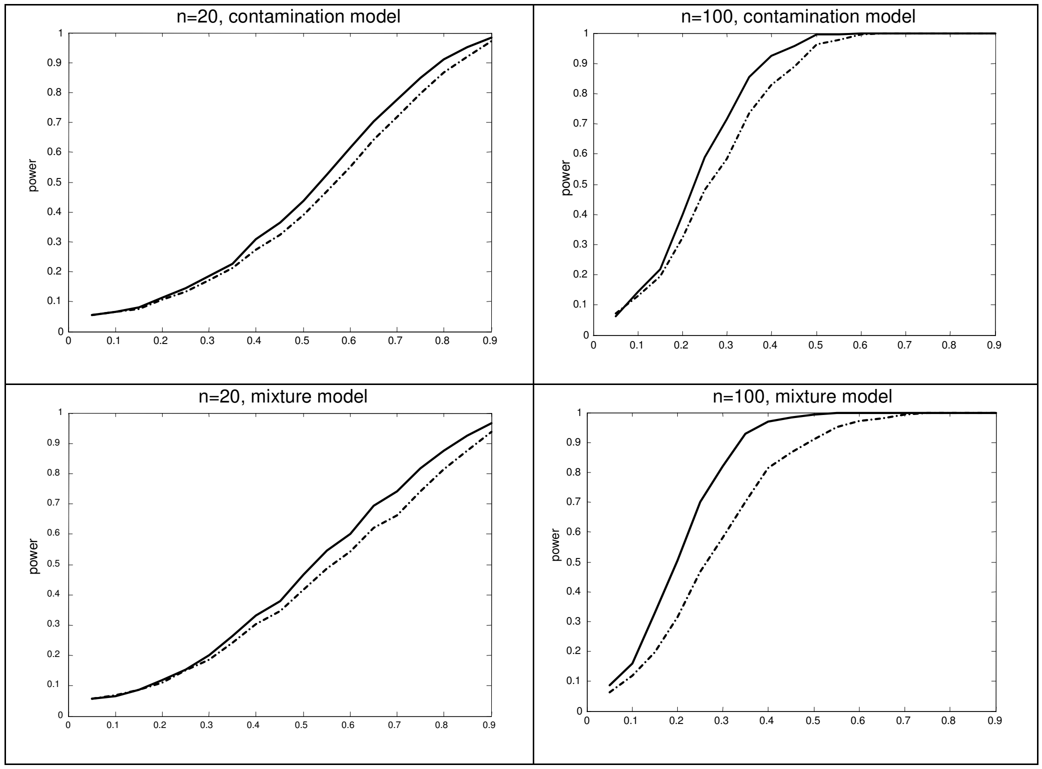

Figure 4.6.5 Power graphs of W^2 and W_1^2 (Student's t), contamination and mixture models.

Again, the W^2 test is more robust and powerful than the W_1^2 test. **It seems the time has come to shift to the new methodology of modified maximum likelihood from the traditional methodology which exclusively engages sample means and variances.**

CHAPTER 5

ILLUSTRATIVE EXAMPLES

In this chapter we provide a few real life and computer generated data sets and analyze them using the methods developed in previous chapters. We test the hypothesis that the population means are equal using the statistics defined in Chapter 4. We also perform hypothesis tests involving the correlation coefficient. The advantage of computer generated data is that the true values of the population parameters are known. One can, therefore, evaluate the accuracy of the estimates obtained by using different methods.

5.1 Examples using simulated data

At first we give two examples using simulated data.

Example 5.1.1

We simulate data in the situation where the marginal and conditional distributions (distributions of X and Y/X, and similarly of U and V/U) are both generalized logistic.

We assume that $b_x = 0.5$, $b_y = 1.0$, $b_u = 1.0$ and $b_v = 0.5$. Also, we assume that the true value of the parameters are:

$\mu_x = 0, \mu_u = 0,\ \mu_y = 0,\ \mu_v = 0, \sigma_1 = 1,\ \sigma_2 = 1$ and $\rho = 0.5$.

Table 5.1.1 Simulated data (GL) for n = 20.

x_i	y_i	x_i	y_i	u_i	v_i	u_i	v_i
2.642	3.186	-1.195	0.822	0.807	-3.007	-2.445	-0.766
0.159	-1.412	-1.049	-0.046	1.268	1.577	-0.324	-1.111
0.214	2.395	-3.578	-1.403	0.382	0.641	3.383	1.101
0.16	-1.254	-6.962	-2.995	-3.234	-2.234	-2.525	-2.192
0.038	1.462	6.476	3.297	2.478	2.829	-0.666	0.975
-3.534	0.591	-5.097	-2.549	-0.894	0.847	1.472	-1.759
0.568	-0.284	-1.629	-2.468	-1.381	-2.465	-3.46	-1.242
1.294	-0.269	-1.878	-2.485	1.094	-1.813	-0.569	-7.238
-3.406	-2.169	0.663	0.465	1.845	-0.84	-0.129	-2.642
-0.355	-2.264	1.181	-0.879	-0.86	-1.178	1.718	-3.956

Table 5.1.2 The MML and LS estimates for the simulated data in Table 5.1.1.

n = 20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
MML	0.694	0.185	-0.07	0.074	1.203	1.067	0.505
LS	0.771	0.234	-0.102	-0.059	1.107	0.961	0.486

It is interesting to see how close the MML and the LS estimates are to one another.

Using the above simulated data, we test the hypothesis:

$$H_0 : \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} \mu_u \\ \mu_v \end{bmatrix} \text{ or equivalently } H_0 : \begin{bmatrix} \mu_x - \mu_u \\ \mu_{y/x} - \mu_{v/u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

To test H_0 we calculate $\hat{T}^2$:

$$\hat{T}^2 = n\left[\hat{\mu}_x - \hat{\mu}_u \quad \hat{\mu}_{y/x} - \hat{\mu}_{v/u}\right] \hat{\Omega}^{-1} \begin{bmatrix} \hat{\mu}_x - \hat{\mu}_u \\ \hat{\mu}_{y/x} - \hat{\mu}_{v/u} \end{bmatrix}$$

$$= 20\begin{bmatrix}0.764 & -0.231\end{bmatrix}\begin{bmatrix}11.05 & 0 \\ 0 & 6.48\end{bmatrix}^{-1}\begin{bmatrix}0.764 \\ -0.231\end{bmatrix}$$

$$= \frac{20(0.764)^2}{11.05} + \frac{20(-0.231)^2}{6.48} = 1.22.$$

Under H_0, $\hat{T}^2$ is distributed (approximately) as chi-square with 2 degrees of freedom. Here, we simulate the p-value of the above test. The simulated p-value is 0.535. The p-value we obtain from chi-square distribution with 2 degrees of freedom is 0.543, which is very close to the simulated value. This p-value is greater than 0.05. Therefore, we do not reject H_0.

Now, using the LSE we calculate $\tilde{T}^2$:

$$\tilde{T}^2 = \frac{20(0.873)^2}{12.1} + \frac{20(-0.075)^2}{6.97} = 1.28.$$

Under H_0, $\tilde{T}^2$ is distributed (approximately) as chi-square with 2 degrees of freedom. The simulated p-value in this case is 0.523. The p-value found by using the chi-square distribution is 0.529; H_0 is not rejected. The two tests are in agreement.

Notice that even for a sample size as small as n = 20, the null distributions of $\hat{T}^2$ and $\tilde{T}^2$ are very close to the asymptotic distribution (chi-square with 2 degrees of freedom).

Table 5.1.3 shows the results for a data of size 100 simulated from the generalized logistic distributions with the same parameters as above. We also give the simulated variances. Notice that the variances of the MML estimators are smaller than the corresponding LS estimators. That illustrates the overall superiority of the MML estimators.

Table 5.1.3 The MML and LS estimators for simulated data (GL), assuming
$\mu_x = 0, \mu_u = 0,\ \mu_y = 0,\ \mu_v = 0, \sigma_1 = 1,\ \sigma_2 = 1$ and $\rho = 0.5$.

n = 100	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
MML	0.186	-0.147	-0.201	-0.037	1.087	1.029	0.511
LS	0.234	-0.152	-0.19	-0.083	1.168	1.022	0.529
Simulated Variance:							
MML	0.051	0.034	0.028	0.041	0.005	0.003	0.003
LS	0.056	0.041	0.028	0.054	0.018	0.008	0.006

Example 5.1.2

Here, we simulate data in the situation where the marginal and conditional distributions (distributions of X and Y/X, and similarly of U and V/U) are both from the Student's t family.
We assume that $r_x = 4$, $r_y = 4$, $r_u = 6$ and $r_v = 6$. Also, we assume that the true value of the parameters are:

$\mu_x = 0, \mu_u = 0,\ \mu_y = 0,\ \mu_v = 0, \sigma_1 = 1,\ \sigma_2 = 1$ and $\rho = 0.2$.

Table 5.1.4 Simulated data (Student's t) for n = 20.

x_i	y_i	x_i	y_i	u_i	v_i	u_i	v_i
0.46	2.682	-0.273	-1.259	1.062	0.733	0.523	0.404
1.54	1.204	0.217	0.067	0.144	-1.105	0.925	0.773
1.509	-2.136	1.004	0.52	1.126	0.566	0.174	2.339
0.438	-0.446	0.088	-1.014	1.115	-0.05	-0.878	0.091
-0.771	0.696	-0.461	0.52	-0.326	0.253	0.624	0.466
0.702	0.579	0.195	-0.079	0.924	0.331	0.36	0.628
0.372	-0.586	-2.854	-1.436	0.293	-1.174	0.434	0.602
-0.743	-0.242	-0.785	-2.586	-0.432	0.901	2.26	1.471
-0.634	-0.574	1.077	1.561	0.779	0.147	0.698	0.217
1.169	0.116	0.163	-0.683	5.607	-1.493	-2.054	-2.371

Table 5.1.5 The MML and LS estimates for the simulated data in Table 5.1.4.

n = 20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
MML	0.168	-0.137	0.571	0.26	0.982	1.047	0.238
LS	0.121	-0.155	0.668	0.186	0.951	0.893	0.158

Using the above simulated data, we test the hypothesis:

$$H_0 : \begin{bmatrix} \mu_x - \mu_u \\ \mu_{y/x} - \mu_{v/u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

To do this, we calculate $\hat{T}^2$:

$$\hat{T}^2 = n\left[\hat{\mu}_x - \hat{\mu}_u \quad \hat{\mu}_{y/x} - \hat{\mu}_{v/u}\right] \hat{\Omega}^{-1} \begin{bmatrix} \hat{\mu}_x - \hat{\mu}_u \\ \hat{\mu}_{y/x} - \hat{\mu}_{v/u} \end{bmatrix}$$

$$= 20\left[-0.403 \quad -0.294\right] \begin{bmatrix} 2.40 & 0 \\ 0 & 2.57 \end{bmatrix}^{-1} \begin{bmatrix} -0.403 \\ -0.294 \end{bmatrix}$$

$$= \frac{20(-0.403)^2}{2.4} + \frac{20(-0.294)^2}{2.57} = 2.03.$$

Under H_0, $\hat{T}^2$ is distributed (approximately) as chi-square with 2 degrees of freedom. The simulated p-value is 0.326, whereas the p-value found using the chi-square distribution is 0.362. Since the p-value is greater than 0.05; we do not reject H_0.

Now using the LSE we calculate $\tilde{T}^2$:

$$\tilde{T}^2 = \frac{20(-0.547)^2}{3.16} + \frac{20(-0.26)^2}{2.72} = 2.39.$$

Under H_0 , $\tilde{T}^2$ is distributed (approximately) as chi-square with 2 degrees of freedom. The simulated p-value here is 0.317 and the p-value calculated by using chi-square distribution with 2 df is 0.303. Again, H_0 is not rejected.

Table 5.1.6 The MML and LS estimators for simulated data (Student's t), assuming $\mu_x = 0, \mu_u = 0, \mu_y = 0, \mu_v = 0, \sigma_1 = 1, \sigma_2 = 1$ and $\rho = 0.2$.

n = 100	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
MML	-0.042	-0.026	0.129	-0.122	1.062	1.034	0.225
LS	-0.079	-0.014	0.202	-0.152	1.014	0.945	0.224
Simulated Variance:							
MML	0.015	0.014	0.014	0.015	0.005	0.004	0.006
LS	0.019	0.015	0.016	0.018	0.007	0.003	0.007

5.2 Examples Using Real Data

Example 5.2.1

The data we use in this example is given in Appendix G (see Beall, 1945). It contains four psychological test scores on 32 Males and 32 Females. The variables in the data are as follows:

y_1 = pictorial inconsistencies. y_3 = tool recognition.

y_2 = paper from board. y_4 = vocabulary.

We choose two of these variables to run our analysis. For illustration, we choose the 1st and 4th variables. Let

X = vocabulary score for males (y_4),

Y = pictorial inconsistency score for males (y_1),

U = vocabulary score for females (y_4), and

V = pictorial inconsistency score for females (y_1).

We first test the hypothesis of bivariate normality using the samples (x_i, y_i) and (u_i, v_i). For the first sample the value of Sürücü test statistic is C_2 = 0.1228 and the critical value is 0.1065 (see Sürücü, 2006). For the second sample the value of Sürücü test statistic is C_2 = 0.2328 and the critical value is 0.1065. In both cases we see that bivariate normality is not applicable. Thus, using the traditional estimators based on normality assumption is not appropriate. We now try to find the underlying distributions, marginal (X or U) as well as the conditional (Y/X and V/U).

We start with the Males and calculate the LSE $\tilde{\theta}$ (= 0.287) and the deviants

$\tilde{e}_{1i} = y_i - 0.287x_i$ ($1 \le i \le 32$).

We construct a Q-Q plot of the ordered deviants $\tilde{e}_{1(i)} = y_{[i]} - 0.287x_{[i]}$ against the quantiles $Q_i = F^{-1}\left(\frac{i}{n+1}\right)$ ($1 \le i \le 32$), F(z) being the CDF of the standard normal distribution. The Q-Q plot is given below.

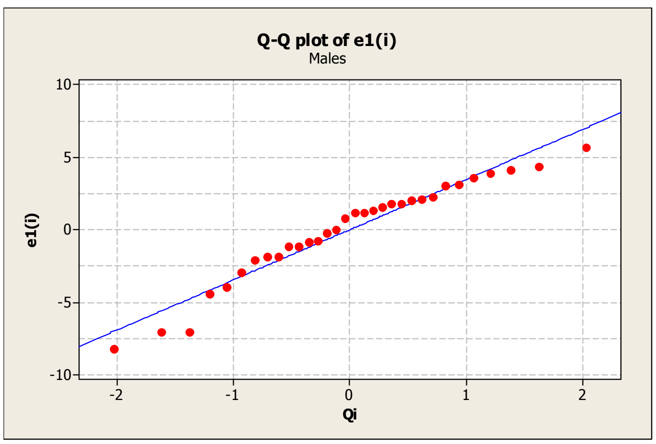

Figure 5.2.1 Normal Q-Q plot of the ordered deviants $\tilde{e}_{1(i)}$ in example 5.2.1.

We notice that the distribution of the errors has a slight negative skewness. The distribution of e_1 is probably generalized logistic (see equation (2.4.2)). To find the value of b_y, we choose the value of b_y which maximizes the likelihood function of the e_{1i}'s. In this case it is b_y = 1.0. Although the distribution looks a little skewed, the value of b_y which maximizes the likelihood function is 1.0, i.e., the most plausible underlying distribution is logistic.

To find the distribution of the x's, a Q-Q plot of the order statistics $x_{(i)}$'s against Q_i's is given below.

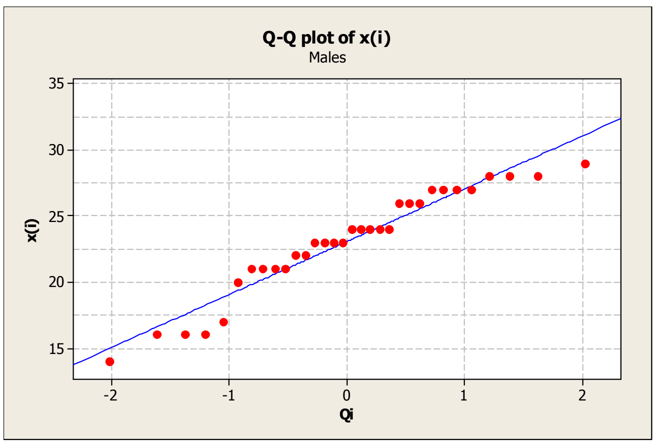

Figure 5.2.2 Normal Q-Q plot of the order statistics $x_{(i)}$'s in example 5.2.1.

Again the distribution of the x's is slightly negatively skewed. The generalized logistic (see equation (2.4.1)) fits the data well. The value of b_x = 1.0 is the one which maximizes the likelihood function of the x_i's.

We do the same for the Females and calculate the LSE $\tilde{\theta}$ (= 0.473) and the deviants

$$\tilde{e}_{2i} = v_i - 0.473u_i \quad (1 \le i \le 32).$$

We construct a Q-Q plot of the ordered deviants $\tilde{e}_{2(i)} = v_{[i]} - 0.473u_{[i]}$ against the Q_i ($1 \le i \le 32$). The Q-Q plot is given below.

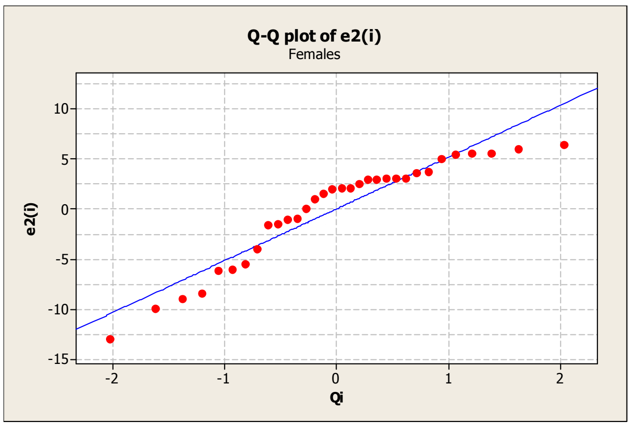

Figure 5.2.3 Normal Q-Q plot of the ordered deviants $\tilde{e}_{2(i)}$ in example 5.2.1.

Clearly, the distribution of the errors is negatively skewed. Calculating the skewness of the residuals in this case, we get -0.8996. The Generalized logistic (see equation (2.4.4)) with shape parameter b_v = 0.5 fits the data well.

We also graph a Q-Q plot of the order statistics $u_{(i)}$'s against Q_i's and again notice that the Generalized logistic (see equation (2.4.3)) with shape parameter $b_u = 0.5$ fits the data well. The Q-Q plot is given in Figure 5.2.4.

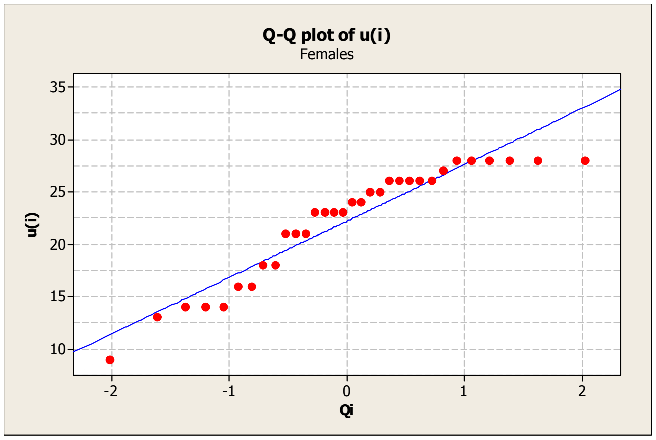

Figure 5.2.4 Normal Q-Q plot of the order statistics $u_{(i)}$'s in example 5.2.1.

We calculate the MML and LS estimates from the data by using the results given in section 2.4. We get the following results:

Table 5.2.1 The MML and LS estimates for the "psychology" data.

n = 32	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
MML	23.34	16.23	25.69	14.53	2.29	1.24	0.32
LS	23.06	15.97	25.15	14.52	2.14	1.22	0.35

Notice that the covariance matrices of the two distributions are essentially equal to one another.

Using the "psychology" data, we test the hypothesis:

$$H_0 : \begin{bmatrix} \mu_x - \mu_u \\ \mu_{y/x} - \mu_{v/u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Here,

$$\hat{T}^2 = n\left[\hat{\mu}_x - \hat{\mu}_u \quad \hat{\mu}_{y/x} - \hat{\mu}_{v/u}\right] \hat{\Omega}^{-1} \begin{bmatrix} \hat{\mu}_x - \hat{\mu}_u \\ \hat{\mu}_{y/x} - \hat{\mu}_{v/u} \end{bmatrix}$$

$$= 32\left[-2.35 \quad 2.11\right] \begin{bmatrix} 40.68 & 0 \\ 0 & 10.76 \end{bmatrix}^{-1} \begin{bmatrix} -2.35 \\ 2.11 \end{bmatrix}$$

$$= \frac{32(-2.35)^2}{40.68} + \frac{32(2.11)^2}{10.76} = 17.6.$$

Under H_0, $\hat{T}^2$ is referred to chi-square distribution with 2 degrees of freedom. The simulated p-value is 0.00051 and the p-value calculated using the chi-square distribution is 0.00015. Thus, we reject H_0 categorically since the p-value is very low.

Now, using the LSE we calculate $\tilde{T}^2$:

$$\tilde{T}^2 = \frac{32(-2.09)^2}{45.06} + \frac{32(1.86)^2}{12.98} = 11.6.$$

Under H_0 , $\tilde{T}^2$ is referred to chi-square distribution with 2 degrees of freedom. The simulated p-value here is 0.005 and the p-value from the chi-square distribution is 0.003; H_0 is again rejected. Notice that the test based on $\hat{T}^2$ has a much smaller p-value which illustrates the fact that for non-normal populations, the test based on $\hat{T}^2$ has enormously higher power than that based on $\tilde{T}^2$.

Example 5.2.2

The data we use in this example is the Peter Mullins "children" data given in Appendix G. The data is given in Seber (1984), page 122-123. A sample of 27 children who had an inborn error of metabolism known as transient neonatal tyrosinemia (TNT) were compared with 27 normal children (the control group). The comparison was done by comparing the children's scores on the Illinois Test of Psycho lingual Ability (ITPA). The test has 10 scores. We will choose as an example two of those scores:

x_3 = visual memory, and x_5 = auditory memory.

Let

X = auditory memory score for the control group (normal group),

Y = visual memory score for the control group,

U = auditory memory score for the TNT group, and

V = visual memory score for the TNT group.

Performing Sürücü test of multivariate normality using the samples (x_i, y_i) and (u_i, v_i) we have (see Sürücü, 2006)

Control group: $C_2 = 0.137$, and

TNT group: $C_2 = 0.128$.

Both tests are significant at 0.1 significance level. We thus reject bivariate normality in both cases. We now try to find the underlying distributions, marginal (X or U) and the conditional (Y/X and V/U).

We start with the control group and calculate the LSE $\tilde{\theta}$ (= 0.207) and the deviants

$\tilde{e}_{1i} = y_i - 0.207 x_i \quad (1 \le i \le 27)$.

A Q-Q plot of the ordered deviants $\tilde{e}_{1(i)} = y_{[i]} - 0.207 x_{[i]}$ against the quantiles

$Q_i = F^{-1}\left(\frac{i}{n+1}\right)$ $(1 \le i \le 27)$ is given below.

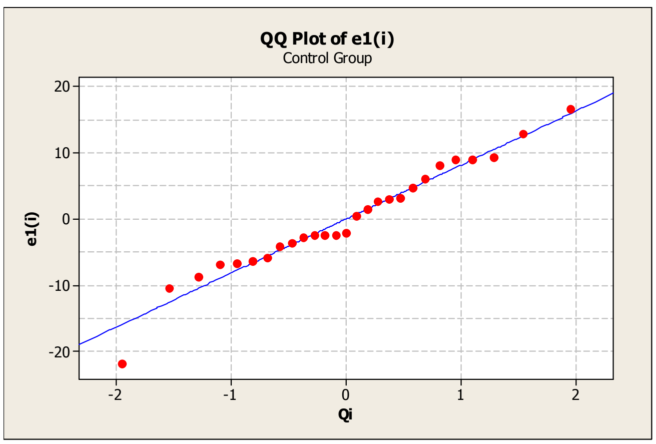

Figure 5.2.5 Normal Q-Q plot of the ordered deviants $\tilde{e}_{1(i)}$ in example 5.2.2.

We notice that the distribution of the errors is close to that of Student's t distribution (see equation (2.5.2)). Choosing the degrees of freedom r_y = 7 provides the best fit for the data.

To find the distribution of the x's, a Q-Q plot of the order statistics $x_{(i)}$'s against Q_i's is given below. Here, the distribution of the *x*'s closely follows a Student's t distribution (see equation (2.5.1)) with r_x = 10.

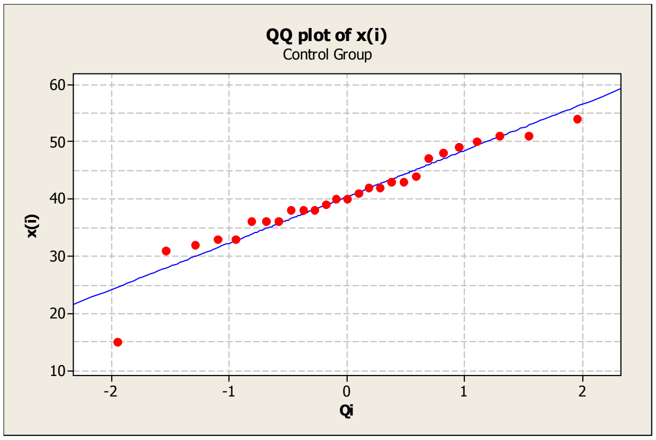

Figure 5.2.6 Normal Q-Q plot of the order statistics $x_{(i)}$'s in example 5.2.2.

We do the same for the TNT group and calculate the LSE $\tilde{\theta}$ (= 0.009) and the deviants

$$\tilde{e}_{2i} = v_i - 0.009 \ u_i \quad (1 \le i \le 27).$$

The Q-Q plot of the ordered deviants is given in Figure 5.2.7. Student's t distribution (see equation (2.5.4)) with r_v = 10 fits the data beautifully well. Also, a Q-Q plot of the ordered statistics $u_{(i)}$'s against Q_i's is given in Figure 5.2.7. Notice that the distribution given in equation (2.5.3) with r_u = 10 fits the data fairly well.

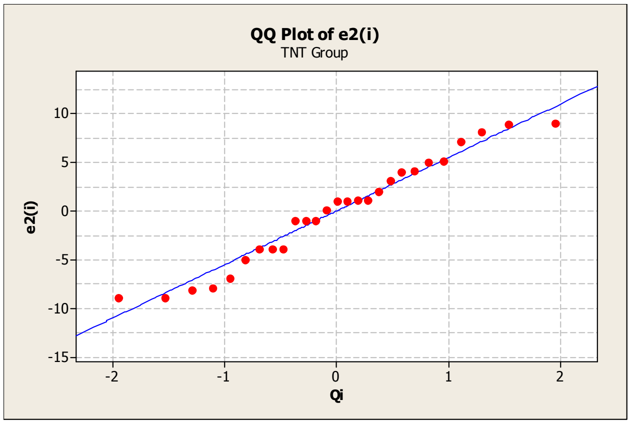

Figure 5.2.7 Normal Q-Q plot of the ordered deviants $\tilde{e}_{2(i)}$ in example 5.2.2.

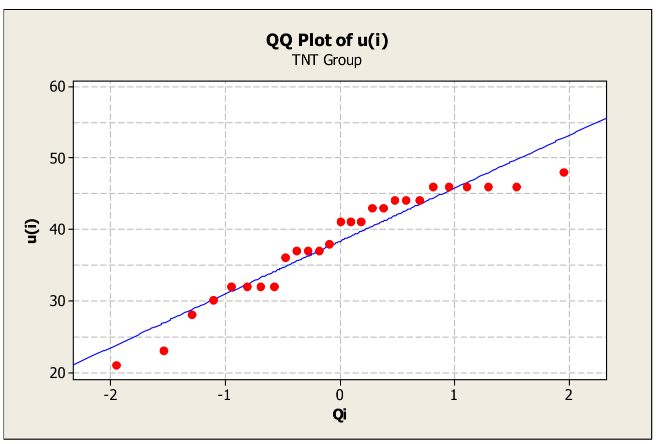

Figure 5.2.8 Normal Q-Q plot of the order statistics $u_{(i)}$'s in example 5.2.2.

We now calculate the MML and LS estimates from the data by using the formulas given in section 2.5. They are given in Table 5.2.2.

Table 5.2.2 The MML and LS estimates for the "Children" data

n = 27	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
MML	40.59	38.31	38.74	35.01	7.28	6.68	0.09
LS	40.37	38.30	38.30	34.96	6.94	6.26	0.12

Again the covariance matrices of the two distributions are essentially equal to one another.

Using the data set, we test the hypothesis:

$$H_0 : \begin{bmatrix} \mu_x - \mu_u \\ \mu_{y/x} - \mu_{v/u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Here,

$$\hat{T}^2 = n\left[\hat{\mu}_x - \hat{\mu}_u \quad \hat{\mu}_{y/x} - \hat{\mu}_{v/u}\right] \hat{\Omega}^{-1} \begin{bmatrix} \hat{\mu}_x - \hat{\mu}_u \\ \hat{\mu}_{y/x} - \hat{\mu}_{v/u} \end{bmatrix}$$

$$= 27\left[1.85 \quad 3.14\right] \begin{bmatrix} 120.88 & 0 \\ 0 & 103.56 \end{bmatrix}^{-1} \begin{bmatrix} 1.85 \\ 3.14 \end{bmatrix}$$

$$= \frac{27(1.85)^2}{120.88} + \frac{27(3.14)^2}{103.56} = 3.33.$$

Under H_0, $\hat{T}^2$ is referred to chi-square distribution with 2 degrees of freedom. The simulated p-value is 0.172 and the p-value calculated using the chi-square distribution with 2 df is 0.189. Thus, we fail to reject H_0 at the 0.05 significance level.

Now using the LSE we calculate $\tilde{T}^2$:

$$\tilde{T}^2 = \frac{27(2.07)^2}{120.46} + \frac{27(3.11)^2}{102.23} = 3.52.$$

Under H_0 , $\tilde{T}^2$ is referred to chi-square distribution with 2 degrees of freedom.

The simulated p-value is 0.182 and the p-value based on the chi-square distribution is 0.172; H_0 is not rejected. The two tests are in agreement.

We notice that the correlation coefficient above is very close to zero. This motivates us to test the hypothesis:

$H_0: \rho = 0$ versus $H_1: \rho \neq 0.$

As given in section 4.5, the test statistic based on the MML estimators is

$t^2 = \hat{\rho}^2 / V$,

where V is the simulated variance of $\hat{\rho}$ when $\rho = 0.$

Using the data above we get

$t^2 = 0.092^2 / 0.0173 = 0.49.$

The null distribution of t^2 is referred to chi-square distribution with one degree of freedom. The simulated p-value is 0.49 and the p-value calculated by using chi-square distribution with 1 df is 0.48. Thus, we fail to reject H_0 at the 0.05 significance level.

Based on the LS estimators, to test H_0, we use the test statistic

$t_1^2 = \tilde{\rho}^2 / V_1$;

V_1 is the simulated variance of $\tilde{\rho}$ when $\rho = 0.$

Using the data we get

$t_1^2 = 0.12^2 / 0.02 = 0.72.$

The null distribution of t_1^2 is referred to the chi-square distribution with one degree of freedom. The simulated p-value in this case is 0.39 and the p-value calculated form the chi-square distribution is 0.40. We again fail to reject H_0.

Example 5.2.3

We use the iris data in this example. The iris data was collected by Anderson (1935) and used by Fisher (1936). It has measurements of sepal length, sepal width, petal length and petal width of 50 iris setosa, 50 iris versicolor, and 50 iris virginica plants. Ironically, the underlying distribution has been assumed to be multivariate normal by many authors including Fisher, but the data is known to be nonnormal (Loonly, 1995; Sürücü 2006). We consider the first two groups, namely, the iris setosa and iris virginica. We consider only two variables, the sepal width and the petal length of each plant. Let

X = Sepal width of the iris setosa plants,
Y = Petal length of the iris setosa plants,
U = Sepal width of the iris virginica plants, and
V = Petal length of the iris virginica plants.

Iris Setosa Plants:

Tiku and Akkaya (2004, Chapter 11) showed that the sepal width for the setosa group can adequately be modeled by the LTS distribution with p = 6. Thus, the distribution family given in equation (2.5.1) is a good fit with r_x =

$2p-1=11$. The Q-Q plot of the order statistics $x_{(i)}$'s against Q_i's is given in Figure 5.2.9.

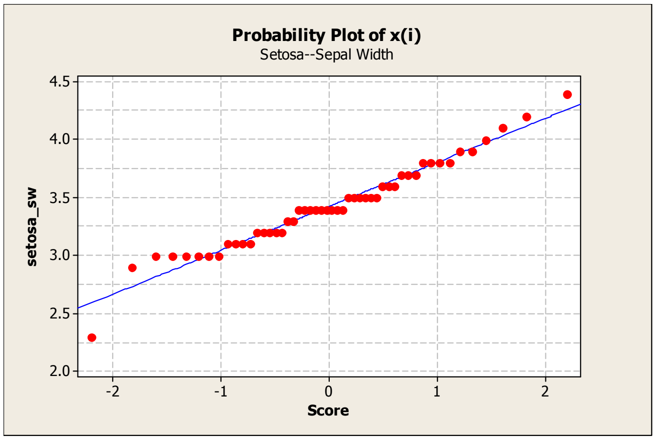

Figure 5.2.9 Normal Q-Q plot of the order statistics $x_{(i)}$'s in example 5.2.3.

To find the distribution of the errors we calculate the LSE $\tilde{\theta}$ (= 0.08) and the deviants

$$\tilde{e}_{1i} = y_i - 0.08x_i \quad (1 \le i \le 50).$$

We construct a Q-Q plot of the ordered deviants $\tilde{e}_{1(i)} = y_{[i]} - 0.08x_{[i]}$ against the quantiles Q_i ($1 \le i \le 50$) of the standard normal distribution. The graph is given in Figure 5.2.10. We notice that the distribution of the errors is close to Student's t distribution (see equation (2.5.2)) with $r_y = 6$.

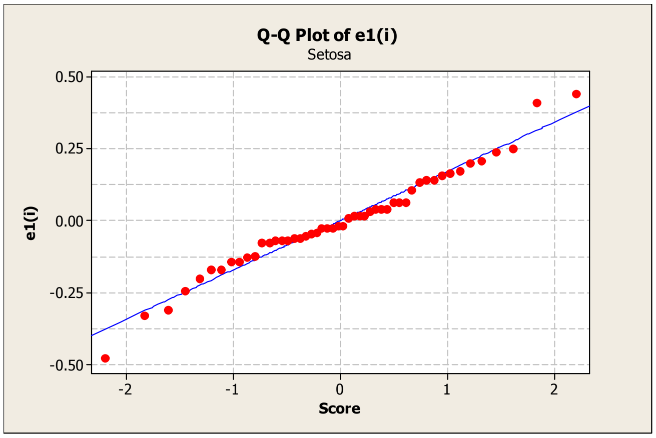

Figure 5.2.10 Normal Q-Q plot of the ordered deviants $\tilde{e}_{1(i)}$ in example 5.2.3.

Iris Virginica Plants:

The Q-Q plot of the ordered statistics $u_{(i)}$'s against Q_i's is given below.

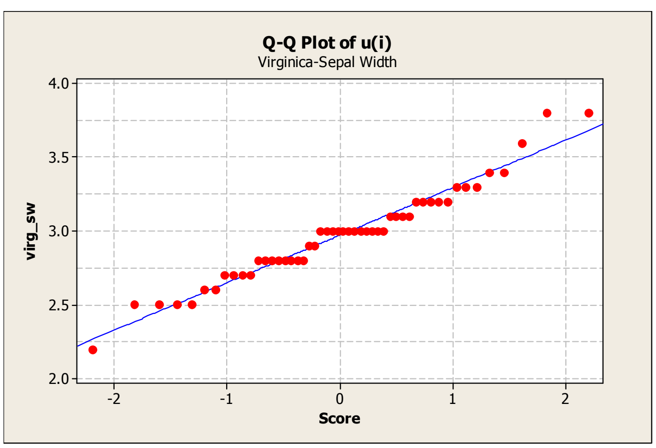

Figure 5.2.11 Normal Q-Q plot of the order statistics $u_{(i)}$'s in example 5.2.3.

We notice again that the distribution given in equation (2.5.3) with $r_u = 11$ fits the data well.

Now, for the sepal width and petal length we calculate the LSE $\tilde{\theta}$ (= 0.67) and the deviants

$$\tilde{e}_{2i} = v_i - 0.67\ u_i \quad (1 \le i \le 50).$$

A Q-Q plot of the ordered deviants $\tilde{e}_{2(i)} = v_{[i]} - 0.67\ u_{[i]}$ against Q_i ($1 \le i \le 50$) is given in Figure 5.2.12. The distribution of the errors closely follows equation (2.5.4) with $r_v = 7$.

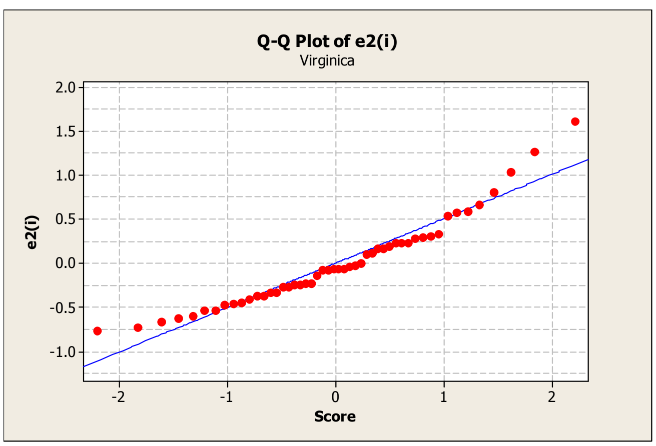

Figure 5.2.12 Normal Q-Q plot of the ordered deviants $\tilde{e}_{2(i)}$ in example 5.2.3.

We calculate the MML and LS estimators from the data using the results given in section 4.6.2. We have the following estimates:

Table 5.2.3 The MML and LS estimates for the iris data.

	μ_x	μ_y	μ_u	μ_v	σ_x	σ_y	σ_u	σ_v	ρ_{yx}	ρ_{vu}
MML	3.42	1.46	2.97	5.50	0.34	0.15	0.29	0.50	0.18	0.47
LS	3.43	1.46	2.97	5.55	0.34	0.14	0.29	0.48	0.19	0.42

Using the iris data set, we test the hypothesis:

$H_0: \rho_{yx} = \rho_{vu}$ vs. $H_1: \rho_{yx} \neq \rho_{vu}$.

As given in section 4.6.3, in order to test this hypothesis, the test statistic based on the MMLE is

$$W^2 = \frac{\left\{0.5\ln\left[\frac{1+\hat{\rho}_{yx}}{1-\hat{\rho}_{yx}}\right] - 0.5\ln\left[\frac{1+\hat{\rho}_{vu}}{1-\hat{\rho}_{vu}}\right]\right\}^2}{2/(n-3)}.$$

Using the data we get

$$W^2 = \frac{\left\{0.5\ln\left[\frac{1+0.183}{1-0.183}\right] - 0.5\ln\left[\frac{1+0.47}{1-0.47}\right]\right\}^2}{2/(50-3)} = 2.5.$$

The null distribution of W^2 is referred to chi-square with one degree of freedom. The simulated p-value is 0.22 and the p-value calculated using the chi-square distribution with 1 df is 0.11. Thus, H_0 is not rejected at the 0.05 significance level.

Based on the LS estimators we have the following test statistic to test H_0

$$W_1^{\,2} = \frac{\left\{0.5\ln\left[\dfrac{1+\tilde{\rho}_{yx}}{1-\tilde{\rho}_{yx}}\right] - 0.5\ln\left[\dfrac{1+\tilde{\rho}_{vu}}{1-\tilde{\rho}_{vu}}\right]\right\}^2}{2/(n-3)}.$$

Using the data we get

$$W_1^{\,2} = \frac{\left\{0.5\ln\left[\dfrac{1+0.194}{1-0.194}\right] - 0.5\ln\left[\dfrac{1+0.421}{1-0.421}\right]\right\}^2}{2/(50-3)} = 1.49.$$

Again, the null distribution of $W_1^{\,2}$ is referred to chi-square with one degree of freedom. The simulated p-value is 0.18 and the p-value using the chi-square distribution is 0.22. Again, H_0 is not rejected.

CHAPTER 6

CONCLUSION

Bivariate data is often assumed to come from a bivariate normal distribution. To estimate parameters in this situation, the least squares method is the best choice. The estimators it yields are equivalent to the MLE and are fully efficient. However, if the distribution is not normal, the LSE loose their efficiency and tend to have a large bias in some situations. Although we can easily adjust for the bias, we must use a method of estimation which gives efficient and robust estimators. In Chapter 1, we summarized three methods of estimation, namely, the maximum likelihood, the modified maximum likelihood, and the least squares method. We discussed the advantages and disadvantages of each method so far as their efficiencies and robustness properties are concerned.

In Chapter 2, we considered two sets of bivariate data coming from a distribution different than the bivariate normal. We considered two specific distributions, the Generalized Logistic and the Student's t. We used the method of modified maximum likelihood to find estimators of all the parameters. We also considered the corresponding LS estimators adjusted for bias. We evaluated the Fisher information matrix and showed that the MMLE are asymptotically fully efficient and considerably more efficient than the LSE.

In Chapter 3, we carried out a simulation study and compared the efficiency and robustness of the MML estimators with those the LS estimators derived in Chapter 2. The MML estimators turned out to have less bias than the LS estimators. They also turned out to be more efficient than the LSE and more robust to the presence of outliers in the data and other data anomalies.

Testing the hypothesis that two population mean vectors are equal is of enormous importance. In Chapter 4, using the MML estimators derived in Chapter 2, we formulated a Hotelling type T^2 statistic for testing the hypothesis that the population mean vectors are equal, and denoted the test statistic by $\hat{T}^2$. We derived the corresponding test statistic based on the LS estimators and denoted it by $\tilde{T}^2$. We derived their asymptotic noncentrality parameters and proved that $\hat{T}^2$ provides a more powerful test than $\tilde{T}^2$ (asymptotically). We simulated the powers of each test and showed that the test based on the MMLE has higher power than the test based on the LSE. This is especially true if there are data anomalies since the LS estimators are adversely affected to a greater extent by the presence of data anomalies. The LS estimators are particularly affected by the presence of outliers in the data. The MML estimators are calculated such that the extreme observations in the sample get small weights. Thus, the influence of the extreme observations is depleted. As a consequence, the MMLE inherit desirable robustness properties. We compared the power of the test statistic $\hat{T}^2$ to the test statistic given by Tiku and Singh (1982), T_D^2, based on censored normal samples. We showed that $\hat{T}^2$ has overall higher power than T_D^2. Interestingly, however, T_D^2 has competing power properties for some distributions besides being much easier to compute. We showed that the MML estimators can be used to find an estimator of Mahalanobis distance; in fact, they give a much better estimator than if we were to use the LS estimators.

In Chapter 4 we also formulated hypothesis tests of the correlation coefficient. We formulated test statistics to test hypotheses of the form H_0: $\rho = 0$ (if the var-cov matrices of both populations are equal) and $\rho_{yx} = \rho_{vu}$ (if the var-cov matrices of both populations are not equal). We showed that the tests based on the MMLE are more powerful than those based on the LSE.

In Chapter 5, we provided examples of real life (and computer generated) data. We used the MML and LS estimators to estimate population parameters. From the data, we calculated $\hat{T}^2$ and $\tilde{T}^2$ statistics and showed how they can be used to test the hypothesis that the population mean vectors are equal. We also considered testing H_0: $\rho = 0$ and $\rho_{yx} = \rho_{vu}$.

To sum up, we showed that using the method of modified maximum likelihood to analyze bivariate data is advantageous. Not only does it give unbiased (almost), efficient and robust estimators, but it also yields more powerful tests for testing statistical hypothesis such as equality of two population mean vectors, or the equality of two correlation coefficients. Although we have only illustrated this for the Generalized Logistic and the Student's t distributions, the MML method can be applied to a wide range of location-scale type distributions. Extension of the methodology developed in this thesis to higher dimensions will indeed be a welcome contribution.

REFERENCES

Abramowitz, M. and Stegun, I.A., *Handbook of Mathematical Functions*, Dover, New York, 1965.

Akkaya, A.D. & Tiku, M.L. (2005a). Time Series AR(1) model for short-tailed distributions, *Statistics*, 39: 117-132.

Akkaya, A.D. & Tiku, M.L. (2005b). Robust estimation and hypothesis testing under short-tailedness and inliers, *Test*, 14: 129-150.

Anderson, E. (1935). The Irises of the Gaspe' Peninsula, *Bulletin of the American Iris Society*, 59: 2-5.

Balakrishnan, N. and Leung, W. Y. (1988). Means, variances and covariances of order statistics, BLUE for the Type I generalized logistic distribution and some applications, *Communications in Statistics – Simulations*, 17, 51- 84

Beall, G. (1945). Approximate Methods in Calculating Discriminant Functions, *Psychometrica*, 10(3): 205-217.

Bhattacharyya, G.K. (1985). The asymptotics of maximum likelihood and related estimators based on type II censored data, *Journal of the American Statistical Association*, 80: 398-404.

Filliben, J. J. (1975). The Probability Plot Correlation Coefficient Test for Normality, *Technometrics*, 17: 111-117.

Fisher, R. A. (1936). The use of Multiple Measurements in Taxonomic Problems, *Ann. Eugen.*, 7: 179-188.

Gayen, A. K. (1951). The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes, *Biometrika*, 38, 219-228.

Islam, M.Q., Tiku, M. L. and Yıldırım, F. (2001). Non-normal Regression, Part 1: Skew distributions. *Communications in Statistics-Theory and Methods*, 30, 993-1020.

Islam, M.Q. & Tiku, M.L. (2004). Multiple regression model under non-normality, *Communications in Statistics-Theory and Methods*, 33: 993-1020.

Islam, M.Q. and Tiku, M. L. (2007). Multiple regression model with stochastic design variables. *Journal of Multivariate Analysis* (under review).

Johnson, R. A. & Wichern, D. W., *Applied Multivariate Statistical Analysis*, 2^{nd} edition, Prentice Hall, NJ, 1988.

Jolicoeur, P. & Mosimann, J. E. (1960). Size and Shape Variation in the Painted Turtle: A principal Component Analysis, *Growth*, 24: 339-354

Kendall, M. G., Stuart, A., *The Advanced Theory of Statistics*, Vol. 2, Charles Griffin, London, 1972.

Lehmann, E. L., *Testing Statistical Hypothesis*, *2^{nd} ed.*, John Wiley, New York, 1986.

Looney, S. W. (1995). How to use Tests for Univariate Normality to Assess Multivariate Normality, *Am. Stat.*, 49: 64-70.

Lubischew, A. A. (1962). On the Use of Discriminant Functions in Taxonomy, *Biometrics*, 18, 455-577.

More, Jorge, Burton, Garbow, and Kenneth Hilstrom, *User Guide for MINPACK-1*, Argonne National Labs Report ANL-80-74, Argonne, Illinois, 1980.

Pearson, E. S. (1959). Note on an approximation to the distribution of non-central χ^2 . *Biometrika*, 46: 346.

Puthenpura S. & Sinha, N.K. (1986). Modified maximum likelihood between the distributions of central and non-central F., *Automatica*, 57: 175-179.

Rencher, A. C., *Methods of Multivariate Analysis*, John Wiley & Sons Inc., New York, 1995.

Sazak, H.S. Tiku, M.L. and Islam, M.Q. (2006). Regression Analysis with a Stochastic Design Variable, *International Statistical Review* 74, 1: 77-88.

Seber, G. A. F., *Multivariate Observations*, John Wiley & Sons Inc., New York, 1984.

Shapiro, S. S., Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples), *Biometrika*, 52: 591-611.

Smith, R. M., Bain, L. J. (1976). Correlation Type Goodness-of-Fit Statistics with Censored Sampling, *Commun. Statis.* A5(2): 119-132.

Sürücü, B. (2006). Goodness-of-Fit Tests for Multivariate Distributions, *Communications in Statistics-Theory and Methods*, 35: 1319-1331.

Tiku, M. L. (1963). Chi-Square Approximations for the distributions of goodness-of-fit statistics U_N^2 and W_N^2 , *Biometrika*, 52: 630-633.

Tiku, M.L (1967). Estimating the mean and standard deviation from a censored normal sample. *Biometrika*, 54: 155-165.

Tiku, M.L., and Singh, M. (1982). Robust Statistics for testing mean vectors of multivariate distributions. *Commun. Stat.-Theory Meth.*, A11(9): 985-1001.

Tiku, M.L., Modified maximum likelihood estimation. *Encyclopedia of Statistical Sciences*, Eds. S. Kotz and N.L. Johnson, Supplement Volume. John Wiley, New York, 1989.

Tiku, M.L. & Akkaya, A.D., *Robust Estimation and Hypothesis Testing*, New Age International Ltd., New Delhi, 2004.

Tiku, M.L and Singh, M. (1982). Robust Statistics for Testing Mean Vectors of Multivariate Distributions, *Commun. Stat.-Theory Meth.*, 11(9): 985-1001.

Tiku, M. L. (1988). Modified Maximum Likelihood Estimation for the Bivariate Normal, *Commun. Stat.-Theory Meth.*, 17(3): 893-910.

Tiku, M.L. and Balakrishnan, N. (1988). Robust Hotelling-Type T^2 Statistic Based on the Modified Maximum Likelihood Estimators, *Commun. Stat.-Theory Meth.*, 17(6): 1789-1810.

Tiku, M.L. and Balakrishnan, N. (1988). Generalization of the Robust Bivariate T^2 Statistic to Multivariate Populations, *Commun. Stat.-Theory Meth.*, 17(6): 3899-3911.

Tiku, M.L. & Suresh, R.P. (1992). A new method of estimation for location and scale parameters. *J. Statist. Plann. Inf.*, 30: 281-292

Tiku, M.L., Wong, W.K., & Bian, G. (1999). Time series models with asymmetric innovations. *Commun. Stat.-Theory Meth.*, 28: 1331-1360.

Tiku, M. L., Wong, W. K., Vaughan, D. C. and Bian, G. (2000). Time series models in non-normal situations: symmetric innovations. *J. Time Series Analysis,* 21, 571-595.

Tiku, M.L, Islam, M.Q. and Selcuk, A.S. (2001). Non-normal Regression: Part II, Skew Distributions, *Commun. Stat.-Theory Meth*., 30(6): 993-1020.

Tiku, M.L., Islam, M.Q., and Sazak, H.S. (2007). Estimation in bivariate non-normal distributions with stochastic variance functions. To appear in *Comput. Statist. Data Anal.*, available online doi:10.1016/j.csda.2007.05.027

Travers, R. M. W. (1939). The use of a Discrimant Function in the Treatment of Psychological Group Differences, *Psychometrica*, 4(1): 25-32.

Vaughn, D.C. and Tiku, M.L. (2000). Estimation and Hypothesis Testing for a Nonnormal Bivariate Distribution with Applications, *Mathematical and Computer Modelling*, 32: 53-67.

Vaughn, D. C. (2002). The Generalized Secant Hyperbolic distribution and its properties. *Communications in Statistics-Theory and Methods*, 31, 219-238.

APPENDIX A

BIAS CORRECTION IN THE LEAST SQUARE ESTIMATORS

A.1 Generalized Logistic

Note that from the properties of the Generalized logistic distribution, we have

$E(Z_1) = \psi(b_x) - \psi(1)$, $Var(Z_1) = \psi'(b_x) + \psi'(1)$,

$E(Z_2) = \psi(b_y) - \psi(1)$, $Var(Z_2) = \psi'(b_y) + \psi'(1)$,

$E(W_1) = \psi(b_u) - \psi(1)$, $Var(W_1) = \psi'(b_u) + \psi'(1)$ and

$E(W_2) = \psi(b_v) - \psi(1)$, and $Var(W_2) = \psi'(b_v) + \psi'(1)$.

Adjusting $\tilde{\mu}_x$, $\tilde{\mu}_u$, $\tilde{\mu}_y$ and $\tilde{\mu}_v$ for the bias

$E(Z_1) = \psi(b_x) - \psi(1)$ implies that

$$\psi(b_x) - \psi(1) = E\left(\frac{X - \mu_x}{\sigma_1}\right).$$

Thus, $E(X) = \mu_x + \sigma_1(\psi(b_x) - \psi(1))$.

In the same manner we get that $E(U) = \mu_u + \sigma_1(\psi(b_u) - \psi(1))$.

Also, $E(Z_2) = \psi(b_y) - \psi(1) = E\left(\dfrac{Y - \mu_y - \theta\ (X - \mu_x)}{\sigma_{2.1}}\right)$.

Thus, $E(Y) + \theta\ \mu_x - \mu_y - \theta\ E(X) = \sigma_{2.1}(\psi(b_y) - \psi(1))$.

This gives $E(Y) = \mu_y + \theta\sigma_1(\psi(b_x) - \psi(1)) + \sigma_{2.1}(\psi(b_y) - \psi(1))$.

In the same manner we get that

$$E(V) = \mu_v + \theta\sigma_1(\psi(b_u) - \psi(1)) + \sigma_{2.1}(\psi(b_v) - \psi(1)) .$$

Using the above equations we can see that the estimators of μ_x, μ_u, μ_y and μ_v are, respectively:

$$\tilde{\mu}_x = \overline{x} - (\psi(b_x) - \psi(1))\tilde{\sigma}_1 ,$$

$$\tilde{\mu}_u = \overline{u} - (\psi(b_u) - \psi(1))\tilde{\sigma}_1 ,$$

$$\tilde{\mu}_y = \overline{y} - \tilde{\theta}\tilde{\sigma}_1(\psi(b_x) - \psi(1)) - \tilde{\sigma}_{2.1}(\psi(b_y) - \psi(1)) \text{ and}$$

$$\tilde{\mu}_v = \overline{v} - \tilde{\theta}\tilde{\sigma}_1(\psi(b_u) - \psi(1)) - \tilde{\sigma}_{2.1}(\psi(b_v) - \psi(1)) .$$

Note that the above estimators are almost unbiased. For example, if σ_1 is known, the estimator $\tilde{\mu}_x = \overline{x} - (\psi(b_x) - \psi(1))\sigma_1$ would be unbiased. However, since σ_1 is unknown here, we are forced to replace it by $\tilde{\sigma}_1$.

Adjusting $\tilde{\mu}_{y/x}$ and $\tilde{\mu}_{v/u}$ for the bias

If we let $p_{1i} = y_i - \theta\, x_i$ and $p_{2i} = v_i - \theta\, u_i$, then

$$E(P_1) = E(e_1) + \mu_{y/x} = \sigma_{2.1}(\psi(b_y) - \psi(1)) + \mu_{y/x} \text{ and}$$

$$E(P_2) = E(e_2) + \mu_{v/u} = \sigma_{2.1}(\psi(b_v) - \psi(1)) + \mu_{v/u} .$$

Therefore we can let

$\tilde{\mu}_{y/x} = \overline{p_1} - (\psi(b_y) - \psi(1))\tilde{\sigma}_{2.1}$, and $\tilde{\mu}_{v/u} = \overline{p_2} - (\psi(b_v) - \psi(1))\tilde{\sigma}_{2.1}$.

Replacing theta in p's by its LS estimator we get

$\tilde{\mu}_{y/x} = \bar{y} - \tilde{\theta}\bar{x} - (\psi(b_y) - \psi(1))\tilde{\sigma}_{2.1}$ and

$\tilde{\mu}_{v/u} = \bar{v} - \tilde{\theta}\bar{u} - (\psi(b_v) - \psi(1))\tilde{\sigma}_{2.1}$.

Adjusting $\tilde{\sigma}_{2.1}$ for the bias

Note that $\sum_{i=1}^{n}\left(p_{1i} - \overline{p_1}\right)^2 = \sum_{i=1}^{n}\left(y_i - \bar{y} - \theta(x_i - \bar{x})\right)^2$ and

$\sum_{i=1}^{n}\left(p_{2i} - \overline{p_2}\right)^2 = \sum_{i=1}^{n}\left(v_i - \bar{v} - \theta(u_i - \bar{u})\right)^2$.

$Var(P_1) = Var(e_1) = \sigma_{2.1}^2(\psi'(b_y) + \psi'(1))$, and

$Var(P_2) = Var(e_2) = \sigma_{2.1}^2(\psi'(b_v) + \psi'(1))$.

$$\mathrm{E}\left[\sum\left(y_i - \bar{y} - \theta(x_i - \bar{x})\right)^2 + \sum\left(v_i - \bar{v} - \theta(u_i - \bar{u})\right)^2\right]$$

$$= (n-2)\sigma_{2.1}^2(\psi'(b_y) + \psi'(1)) + (n-2)\sigma_{2.1}^2(\psi'(b_v) + \psi'(1))$$

$$= (n-2)\sigma_{2.1}^2(\psi'(b_y) + \psi'(b_v) + 2\psi'(1)) .$$

Thus, the LS estimator of $\sigma_{2.1}$ is

$$\tilde{\sigma}_{2.1} = \sqrt{\frac{\left[\sum\left(y_i - \bar{y} - \tilde{\theta}(x_i - \bar{x})\right)^2 + \sum\left(v_i - \bar{v} - \tilde{\theta}(u_i - \bar{u})\right)^2\right]/(n-2)}{\psi'(b_y) + \psi'(b_v) + 2\psi'(1)}} .$$

Adjusting $\tilde{\sigma}_1$ for the bias

Since $Z_1 = \dfrac{X - \mu_x}{\sigma_1}$ and $W_1 = \dfrac{U - \mu_u}{\sigma_1}$, we see that

Var(X) = $\sigma_1^2(\psi'(b_x) + \psi'(1))$ and Var(U)= $\sigma_1^2(\psi'(b_u) + \psi'(1))$.

Now,

$$E\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right) = (n-1)Var(X)$$

$$= (n\text{ -}1)\,\sigma_1^2(\psi'(b_x) + \psi'(1)) .$$

Thus,

$$E\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(u_i - \bar{u})^2\right) = (n-1)\sigma_1^2(\psi'(b_x) + \psi'(1) + \psi'(b_u) + \psi'(1)) .$$

So the LS estimator of σ_1 is

$$\tilde{\sigma}_1 = \sqrt{\frac{s_x^2 + s_u^2}{\psi'(b_x) + \psi'(b_u) + 2\psi'(1)}} .$$

A.2 Student's t Distribution

We will minimize:

1. $E_1 = \sum_{i=1}^{n} e_{1i}^2$ and $E_2 = \sum_{i=1}^{n} e_{2i}^2$

2. $\sum(x_i - \mu_x)^2$ and $\sum(u_i - \mu_u)^2$

Note that since

$Z_1 = \dfrac{X - \mu_x}{\sigma_1} \sim t(r_x)$ then $E(Z_1) = 0$ and $Var(Z_1) = \dfrac{r_x}{r_x - 2} = c_x$.

$Z_2 = \sqrt{c_1(x)}\dfrac{Y - \mu_{y/x} - \theta\, x}{\sigma_{2.1}} \sim t(r_y)$ then

$E(Z_2) = 0$ and $Var(Z_2) = \dfrac{r_y}{r_y - 2} = c_y$.

$W_1 = \dfrac{U - \mu_u}{\sigma_1} \sim t(r_u)$ then

$E(W_1) = 0$ and $Var(W_1) = \dfrac{r_u}{r_u - 2} = c_u$.

$W_2 = \sqrt{c_2(u)}\dfrac{V - \mu_{v/u} - \theta\, u}{\sigma_{2.1}} \sim t(r_v)$ then

$E(W_2) = 0$ and $Var(W_2) = \dfrac{r_v}{r_v - 2} = c_v$.

Finding the LS estimators of $\mu_{y/x}$ and $\mu_{v/u}$

Differentiating E_1 with respect to $\mu_{y/x}$ we get

$$\frac{\partial E_1}{\partial \mu_{y/x}} = -2\sum_{i=1}^{n} c_{1i}(y_i - \mu_{y/x} - \theta\, x_i) = 0,$$

and we get $\tilde{\mu}_{y/x} = \bar{y}_{(.)} - \tilde{\theta}\bar{x}_{(.)}$.

The same can be done to find the LS estimator of $\mu_{v/u}$.

Adjusting $\tilde{\sigma}_1$ for the bias

Note that, $E(s_x^2) = \dfrac{r_x}{r_x - 2}\sigma_1^2$ and $E(s_u^2) = \dfrac{r_u}{r_u - 2}\sigma_1^2$.

Thus, $E(s_x^2 + s_u^2) = \sigma_1^2 \left[\dfrac{r_x}{r_x - 2} + \dfrac{r_u}{r_u - 2} \right]$.

Therefore, $\tilde{\sigma}_1$ is taken as

$$\tilde{\sigma}_1 = \sqrt{\frac{s_x^2 + s_u^2}{\dfrac{r_x}{r_x - 2} + \dfrac{r_u}{r_u - 2}}}\,.$$

Adjusting $\tilde{\sigma}_{2.1}$ for the bias

$$\frac{1}{(n-2)} \min \left[\sum_{i=1}^{n} e_{1i}^2 + \sum_{i=1}^{n} e_{2i}^2 \right] =$$

$$\frac{\left\lfloor \sum c_{1i} \left(y_i - \bar{y}_{(.)} - \tilde{\theta}(x_i - \bar{x}_{(.)}) \right)^2 + \sum c_{2i} \left(v_i - \bar{v}_{(.)} - \tilde{\theta}(u_i - \bar{u}_{(.)}) \right)^2 \right\rfloor}{(n-2)}$$

and

$$E\left[\sum c_{1i} \left(y_i - \bar{y}_{(.)} - \theta(x_i - \bar{x}_{(.)}) \right)^2 + \sum c_{2i} \left(v_i - \bar{v}_{(.)} - \theta(u_i - \bar{u}_{(.)}) \right)^2 \right]$$

$$= (n-2) \left[\frac{r_y}{r_y - 2} + \frac{r_v}{r_v - 2} \right] \sigma_{2.1}^2 .$$

So,

$$\tilde{\sigma}_{2.1} = \sqrt{\frac{\left[\sum c_{1i}\left(y_i - \bar{y}_{(.)} - \tilde{\theta}(x_i - \bar{x}_{(.)})\right)^2 + \sum c_{2i}\left(v_i - \bar{v}_{(.)} - \tilde{\theta}(u_i - \bar{u}_{(.)})\right)^2\right]}{(n-2)\left[\frac{r_y}{r_y - 2} + \frac{r_v}{r_v - 2}\right]}}.$$

Finding an estimator of θ

We differentiate E = $E_1 +\ E_2$ with respect to θ and we get

$$\frac{\partial E}{\partial \theta} = -2\sum_{i=1}^{n} c_{1i}(y_i - \mu_{y/x} - \theta x_i)x_i - 2\sum_{i=1}^{n} c_{2i}(v_i - \mu_{v/u} - \theta u_i)u_i = 0\,.$$

Replacing $\mu_{y/x}$ and $\mu_{v/u}$ by their LS estimators and simplifying we get

$$\tilde{\theta} = \frac{\sum_{i=1}^{n} c_{1i}(x_i - \bar{x}_{(.)})y_i + \sum_{i=1}^{n} c_{2i}(u_i - \bar{u}_{(.)})v_i}{\sum_{i=1}^{n} c_{1i}(x_i - \bar{x}_{(.)})^2 + \sum_{i=1}^{n} c_{2i}(u_i - \bar{u}_{(.)})^2}.$$

APPENDIX B

WEIGHTED LEAST SQUARE ESTIMATORS

B.1 Generalized Logistic

Let E $= \frac{1}{Var(e_1)} \sum_{i=1}^{n} \left(e_{1i} - E(e_1)\right)^2 + \frac{1}{Var(e_2)} \sum_{i=1}^{n} \left(e_{2i} - E(e_2)\right)^2$.

If we differentiate E with respect to θ we get:

$$\frac{\partial E}{\partial \theta} = \frac{1}{\sigma_{2.1}^2} \left[\frac{2}{c_y} \sum_{i=1}^{n} (y_i - \mu_{y/x} - \theta\, x_i - (\psi(b_y) - \psi(1))\sigma_{2.1})(x_i - \mu_x) \right.$$

$$\left. + \frac{2}{c_v} \sum_{i=1}^{n} (v_i - \mu_{v/u} - \theta\, u_i - (\psi(b_v) - \psi(1))\sigma_{2.1})(u_i - \mu_u) \right] = 0.$$

Replacing the parameters $\mu_{y/x}, \mu_{v/u}, \theta$ and $\sigma_{2.1}$ by $\tilde{\mu}_{y/x}, \tilde{\mu}_{v/u}, \tilde{\theta}$ and $\tilde{\sigma}_{2.1}$ respectively the above equation reduces to:

$$\frac{1}{c_y} \sum_{i=1}^{n} \left(y_i - \bar{y} - \tilde{\theta}\, (x_i - \bar{x})\right)\left(x_i - \bar{x} + \psi(b_x) - \psi(1))\tilde{\sigma}_1\right)$$

$$+ \frac{1}{c_v} \sum_{i=1}^{n} \left(v_i - \bar{v} - \tilde{\theta}\, (u_i - \bar{u})\right)\left(u_i - \bar{u} + \psi(b_u) - \psi(1))\tilde{\sigma}_1\right) = 0.$$

Multiplying out the parenthesis we get

$$\frac{1}{c_y}\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})-\frac{\tilde{\theta}}{c_y}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{1}{c_v}\sum_{i=1}^{n}(v_i-\bar{v})(u_i-\bar{u})-\frac{\tilde{\theta}}{c_v}\sum_{i=1}^{n}(u_i-\bar{u})^2$$
$$=0.$$

Therefore,

$$\tilde{\theta}\left[\frac{1}{c_y}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{1}{c_v}\sum_{i=1}^{n}(u_i-\bar{u})^2\right]$$
$$=\frac{1}{c_y}\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})+\frac{1}{c_v}\sum_{i=1}^{n}(v_i-\bar{v})(u_i-\bar{u})$$

This gives

$$\tilde{\theta}_w=\frac{\frac{1}{c_y}\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})+\frac{1}{c_v}\sum_{i=1}^{n}(v_i-\bar{v})(u_i-\bar{u})}{\frac{1}{c_y}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{1}{c_v}\sum_{i=1}^{n}(u_i-\bar{u})^2} \text{ or}$$

$$\tilde{\theta}_w=\frac{\frac{1}{c_y}\sum_{i=1}^{n}(x_i-\bar{x})y_i+\frac{1}{c_v}\sum_{i=1}^{n}(u_i-\bar{u})v_i}{\frac{1}{c_y}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{1}{c_v}\sum_{i=1}^{n}(u_i-\bar{u})^2}.$$

Finding the WLS estimator of σ_1

$$\tilde{\sigma}_{1w}^2=\frac{1}{(n-1)}\min\left(\frac{1}{c_y}\sum_{i=1}^{n}(x_i-\mu_x)^2+\frac{1}{c_v}\sum_{i=1}^{n}(u_i-\mu_u)^2\right),$$

means that $\tilde{\sigma}_{1w}^2=\frac{1}{(n-1)}\left(\frac{1}{c_y}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2+\frac{1}{c_v}\sum_{i=1}^{n}\left(u_i-\bar{u}\right)^2\right)=\frac{s_x^2}{c_y}+\frac{s_u^2}{c_v}.$

Now, $E(s_x^2) = (\psi'(b_x) + \psi'(1))\sigma_1^2 = c_x\sigma_1^2$ and

$E(s_u^2) = (\psi'(b_u) + \psi'(1))\sigma_1^2 = c_u\sigma_1^2$.

Which implies that the expected value of the above estimator is:

$$E(\tilde{\sigma}_{1w}^2) = E\left(\frac{s_x^2}{c_y} + \frac{s_u^2}{c_v}\right) = \sigma_1^2\left(\frac{c_x}{c_y} + \frac{c_u}{c_v}\right).$$

Thus, we choose the following WLS estimator of σ_1 :

$$\tilde{\sigma}_{1w} = \sqrt{\frac{\frac{s_x^2}{c_y} + \frac{s_u^2}{c_v}}{\frac{c_x}{c_y} + \frac{c_u}{c_v}}}$$

Finding the WLS estimator of $\sigma_{2.1}^2$

Note that

$$\mathrm{E}\left[\frac{1}{c_y}\sum\left(y_i - \bar{y} - \tilde{\theta}(x_i - \bar{x})\right)^2 + \frac{1}{c_v}\sum\left(v_i - \bar{v} - \tilde{\theta}(u_i - \bar{u})\right)^2\right] = 2\,(n\text{ -}2)\,\sigma_{2.1}^2.$$

Thus, $$\tilde{\sigma}_{2.1w} = \sqrt{\frac{\left[\frac{1}{c_y}\sum\left(y_i - \bar{y} - \tilde{\theta}(x_i - \bar{x})\right)^2 + \frac{1}{c_v}\sum\left(v_i - \bar{v} - \tilde{\theta}(u_i - \bar{u})\right)^2\right]}{2(n-2)}}.$$

B. 2 Student's t

The derivations are similar and will not be included here.

APPENDIX C

INFORMATION MATRIX

C.1 Generalized Logistic

Let Z ~ GL(b) and let

$g(z) = \dfrac{e^{-z}}{1+e^{-z}}$ and $f(z) = \dfrac{e^{-z}}{(1+e^{-z})^2}$.

Note that $g'(z) = -f(z)$.

Now using the Generalized Logistic p.d.f. we can show that

$E(Z) = \psi(b) - \psi(1)$,

$Var(Z) = \psi'(b) + \psi'(1)$,

$E(g(Z)) = (b+1)^{-1}$,

$E(Zg(Z)) = (b+1)^{-1}(\psi(b) - \psi(2))$,

$E(f(Z)) = b(b+1)^{-1}(b+2)^{-1}$,

$E(Zf(Z)) = b(b+1)^{-1}(b+2)^{-1}(\psi(b+1) - \psi(2))$ and

$E(Z^2 f(Z)) = b(b+1)^{-1}(b+2)^{-1}\left[\psi'(b+1) + \psi'(2) + (\psi(b+1) - \psi(2))^2\right]$.

We will only illustrate the procedure for a couple of the elements of the information matrix. The rest are derived in the same way.

The elements of the first fisher information matrix I($\mu_x, \mu_u, \sigma_1, \mu_{y/x}$, $\mu_{v/u}, \sigma_{2.1}$, θ) are derived as follows:

1. $\frac{\partial^2 \ln L}{\partial \mu_x^2} = -\frac{(b_x+1)}{\sigma_1^2}\sum_{i=1}^{n} f(z_{1i})$. Therefore, we get

$$E\left(-\frac{\partial^2 \ln L}{\partial \mu_x^2}\right) = \frac{n(b_x+1)}{\sigma_1^2} b_x (b_x+1)^{-1}(b_x+2)^{-1} = \frac{nb_x}{\sigma_1^2 (b_x+2)}.$$

Thus,

$$I_{\mu_x \mu_x} = \frac{n}{\sigma_1^2}\left[\frac{b_x}{b_x+2}\right].$$

2. $\frac{\partial^2 \ln L}{\partial \mu_x \partial \sigma_1} = -\frac{n}{\sigma_1^2} + \frac{(b_x+1)}{\sigma_1^2}\sum_{i=1}^{n} g(z_{1i}) - \frac{(b_x+1)}{\sigma_1^2}\sum_{i=1}^{n} z_{1i} f(z_{1i})$. This implies that

$$E\left(-\frac{\partial^2 \ln L}{\partial \mu_x \partial \sigma_1}\right)$$

$$= \frac{n}{\sigma_1^2} - \frac{n(b_x+1)}{\sigma_1^2}(b_x+1)^{-1} + \frac{n(b_x+1)}{\sigma_1^2} b_x (b_x+1)^{-1}(b_x+2)^{-1}(\psi(b_x+1) - \psi(2)).$$

Thus,

$$I_{\mu_x \sigma_1} = E\left(-\frac{\partial^2 \ln L}{\partial \mu_x \partial \sigma_1}\right) = \frac{n}{\sigma_1^2}\left[\frac{b_x}{(b_x+2)}\left(\psi(b_x+1) - \psi(2)\right)\right].$$

C.2 Student's t Distribution

In Chapter 2, we derived the information matrices assuming that $c_{1i} = c_{2i} = 1$ for all i.

Now if Z ~ t(r) then

E(Z) = 0 and $Var(Z) = \frac{r}{r-2}$.

Let $g(z) = \frac{z}{\left(1+\frac{z^2}{r}\right)}$ and let $f(z) = \frac{1-\frac{z^2}{r}}{\left(1+\frac{z^2}{r}\right)^2}$.

Note that $g'(z) = f(z)$.

Using the Student's t p.d.f. we can show the following relations:

$E(g(Z)) = 0$,

$E(Zg(Z)) = \frac{r}{r+1}$,

$E(f(Z)) = \frac{r}{r+3}$,

$E(Zf(Z)) = 0$ and

$E(Z^2 f(Z)) = \frac{r(r-3)}{(r+1)(r+3)}$.

Let us show the procedure of finding the elements of the information matrix by illustrating with only a few elements of the first information matrix I.

1. $\frac{\partial^2 \ln L}{\partial \mu_x^2} = -\frac{(r_x+1)}{r_x \sigma_1^2} \sum_{i=1}^{n} f(z_{1i})$. Therefore, we get

$$E\left(-\frac{\partial^2 \ln L}{\partial \mu_x^2}\right) = \frac{n(r_x+1)}{r_x \sigma_1^2} \frac{r_x}{(r_x+3)} = \frac{n}{\sigma_1^2} \frac{(r_x+1)}{(r_x+3)}.$$

So, $I_{\mu_x \mu_x} = \frac{n}{\sigma_1^2}\left[\frac{r_x+1}{r_x+3}\right]$.

2. $\frac{\partial^2 \ln L}{\partial \mu_x \partial \sigma_1} = -\frac{(r_x+1)}{r_x \sigma_1^2}\left[\sum_{i=1}^{n} z_{1i} f(z_{1i}) + \sum_{i=1}^{n} g(z_{1i})\right]$.

$$E\left(-\frac{\partial^2 \ln L}{\partial\mu_x \partial\sigma_1}\right)=0 \text{ since } E(Z_1 f(Z_1))=0 \text{ and } E(g(Z_1))=0.$$

So we get that $I_{\mu_x\sigma_1} = 0$

3. $$\frac{\partial^2 \ln L}{\partial\theta^2} = -\frac{(r_y+1)}{r_y\sigma_{2.1}^2}\sum_{i=1}^{n} x_i^2 f(z_{2i}) - \frac{(r_v+1)}{r_v\sigma_{2.1}^2}\sum_{i=1}^{n} u_i^2 f(w_{2i}).$$

Writing x_i^2 in terms of z_{1i} and u_i^2 in terms of w_{1i} we get

$$\frac{\partial^2 \ln L}{\partial\theta^2} = -\frac{(r_y+1)}{r_y\sigma_{2.1}^2}\left[\sigma_1^2\sum_{i=1}^{n} z_{1i}^2 f(z_{2i}) + 2\mu_x\sigma_1\sum_{i=1}^{n} z_{1i} f(z_{2i}) + \mu_x^2\sum_{i=1}^{n} f(z_{2i})\right]$$
$$-\frac{(r_v+1)}{r_v\sigma_{2.1}^2}\left[\sigma_1^2\sum_{i=1}^{n} w_{1i}^2 f(w_{2i}) + 2\mu_u\sigma_1\sum_{i=1}^{n} w_{1i} f(w_{2i}) + \mu_u^2\sum_{i=1}^{n} f(w_{2i})\right].$$

Since $E(Z_1^2) = Var(Z_1) + (E(Z_1))^2 = \dfrac{r_x}{r_x-2} + 0$ and similarly

$E(W_1^2) = Var(W_1) + (E(W_1))^2 = \dfrac{r_u}{r_u-2}$, we have

$$E\left(-\frac{\partial^2 \ln L}{\partial\theta^2}\right) = \frac{n}{\sigma_{2.1}^2}\left\{\frac{(r_y+1)}{r_y}\left[\sigma_1^2\frac{r_x}{(r_x-2)}\frac{r_y}{(r_y+3)} + \mu_x^2\frac{r_y}{(r_y+3)}\right]\right.$$
$$\left.+\frac{(r_v+1)}{r_v}\left[\sigma_1^2\frac{r_u}{(r_u-2)}\frac{r_v}{(r_v+3)} + \mu_u^2\frac{r_v}{(r_v+3)}\right]\right\}.$$

Simplifying we get that

$$I_{\theta\theta} = \frac{n}{\sigma_{2.1}^2}\left\{\frac{(r_y+1)}{(r_y+3)}\frac{\left[r_x(\sigma_1^2+\mu_x^2)-2\mu_x^2\right]}{(r_x-2)} + \frac{(r_v+1)}{(r_v+3)}\frac{\left[r_u(\sigma_1^2+\mu_u^2)-2\mu_u^2\right]}{(r_u-2)}\right\}.$$

APPENDIX D

ADDITIONAL SIMULATION RESULTS

Table D.1

ρ =0.2, Outlier Model (a), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.002	0.011	-0.011	-0.011	1.47	1.129	0.257
E_LS	-0.001	0.008	-0.017	-0.014	1.517	1.043	0.285
V_MML	0.228	0.158	0.229	0.156	0.094	0.022	0.02
V_LS	0.402	0.172	0.413	0.173	0.162	0.021	0.026
RE (MML/LS)	57	91	56	90	58	104	76
n=30							
E_MML	0.003	0	0.005	-0.009	1.43	1.087	0.261
E_LS	0.001	-0.002	0.008	-0.011	1.546	1.039	0.294
V_MML	0.148	0.106	0.149	0.098	0.054	0.013	0.012
V_LS	0.277	0.119	0.279	0.112	0.117	0.014	0.017
RE (MML/LS)	54	89	53	87	46	97	73
n=100							
E_MML	-0.001	-0.003	0.005	-0.006	1.353	1.032	0.262
E_LS	-0.003	-0.003	0.001	-0.008	1.565	1.03	0.304
V_MML	0.042	0.033	0.04	0.031	0.012	0.004	0.003
V_LS	0.083	0.036	0.074	0.035	0.039	0.004	0.005
RE (MML/LS)	51	90	54	88	30	87	67

Table D.2

ρ =0.9, Outlier Model (a), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.002	0.006	-0.011	-0.014	1.47	1.411	0.933
E_LS	-0.001	0.003	-0.017	-0.02	1.517	1.438	0.943
V_MML	0.228	0.212	0.229	0.215	0.094	0.072	0.001
V_LS	0.402	0.354	0.413	0.363	0.162	0.125	0.001
RE (MML/LS)	57	60	56	59	58	58	100
n=30							
E_MML	0.003	0.002	0.005	0	1.43	1.37	0.937
E_LS	0.001	0	0.008	0.001	1.546	1.461	0.948
V_MML	0.148	0.142	0.149	0.138	0.054	0.041	0.001
V_LS	0.277	0.248	0.279	0.246	0.117	0.09	0.001
RE (MML/LS)	54	57	53	56	46	46	98
n=100							
E_MML	-0.001	-0.002	0.005	0.001	1.353	1.296	0.939
E_LS	-0.003	-0.004	0.001	-0.002	1.565	1.475	0.953
V_MML	0.042	0.039	0.04	0.039	0.012	0.009	0
V_LS	0.083	0.072	0.074	0.067	0.039	0.03	0
RE (MML/LS)	51	55	54	58	30	31	86

Table D.3

ρ =0.2, Outlier Model (b), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.002	0.011	-0.009	-0.014	1.05	1.538	0.138
E_LS	0.002	0.008	-0.009	-0.018	0.987	1.539	0.131
V_MML	0.153	0.226	0.153	0.228	0.02	0.098	0.019
V_LS	0.166	0.386	0.165	0.396	0.019	0.163	0.025
RE (MML/LS)	92	59	93	58	104	60	73
n=30							
E_MML	0.002	-0.002	0.002	-0.011	1.032	1.468	0.141
E_LS	0.003	-0.009	0.002	-0.016	0.99	1.555	0.13
V_MML	0.102	0.148	0.103	0.136	0.013	0.056	0.012
V_LS	0.11	0.269	0.11	0.26	0.013	0.119	0.017
RE (MML/LS)	93	55	93	52	97	47	71
n=100							
E_MML	-0.001	-0.003	0.006	-0.007	1.009	1.351	0.15
E_LS	-0.003	0.001	0.004	-0.014	0.997	1.546	0.131
V_MML	0.03	0.044	0.03	0.041	0.004	0.012	0.003
V_LS	0.033	0.083	0.031	0.082	0.004	0.037	0.005
RE (MML/LS)	93	52	96	51	89	32	72

Table D.4

ρ =0.9, Outlier Model (b), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.002	0.006	-0.009	-0.014	1.05	1.167	0.808
E_LS	0.002	0.005	-0.009	-0.016	0.987	1.124	0.79
V_MML	0.153	0.166	0.153	0.167	0.02	0.023	0.006
V_LS	0.166	0.207	0.165	0.207	0.019	0.031	0.009
RE (MML/LS)	92	80	93	81	104	76	60
n=30							
E_MML	0.002	0.001	0.002	-0.003	1.032	1.134	0.817
E_LS	0.003	-0.001	0.002	-0.005	0.99	1.13	0.79
V_MML	0.102	0.113	0.103	0.108	0.013	0.015	0.003
V_LS	0.11	0.142	0.11	0.139	0.013	0.022	0.006
RE (MML/LS)	93	80	93	78	97	67	52
n=100							
E_MML	-0.001	-0.003	0.006	0.001	1.009	1.086	0.836
E_LS	-0.003	-0.002	0.004	-0.003	0.997	1.129	0.795
V_MML	0.03	0.033	0.03	0.034	0.004	0.004	0.001
V_LS	0.033	0.042	0.031	0.044	0.004	0.006	0.002
RE (MML/LS)	93	77	96	77	89	56	40

Table D.5

ρ =0.2, Outlier Model (c), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.002	0.011	-0.011	-0.015	1.47	1.551	0.182
E_LS	-0.001	0.007	-0.017	-0.019	1.517	1.553	0.185
V_MML	0.228	0.231	0.229	0.234	0.094	0.101	0.052
V_LS	0.402	0.394	0.413	0.403	0.162	0.165	0.077
RE (MML/LS)	57	59	56	58	58	61	67
n=30							
E_MML	0.003	-0.003	0.005	-0.01	1.43	1.482	0.188
E_LS	0.001	-0.009	0.008	-0.015	1.546	1.571	0.191
V_MML	0.148	0.151	0.149	0.139	0.054	0.058	0.034
V_LS	0.277	0.276	0.279	0.268	0.117	0.122	0.055
RE (MML/LS)	54	55	53	52	46	48	62
n=100							
E_MML	-0.001	-0.003	0.005	-0.007	1.353	1.364	0.198
E_LS	-0.003	0.001	0.001	-0.014	1.565	1.565	0.199
V_MML	0.042	0.044	0.04	0.042	0.012	0.012	0.01
V_LS	0.083	0.085	0.074	0.083	0.039	0.038	0.018
RE (MML/LS)	51	52	54	50	30	32	56

Table D.6

ρ =0.9, Outlier Model (c), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.002	0.007	-0.011	-0.016	1.47	1.486	0.881
E_LS	-0.001	0.003	-0.017	-0.022	1.517	1.525	0.882
V_MML	0.228	0.226	0.229	0.228	0.094	0.086	0.004
V_LS	0.402	0.394	0.413	0.403	0.162	0.148	0.007
RE (MML/LS)	57	57	56	57	58	58	60
n=30							
E_MML	0.003	0.001	0.005	0	1.43	1.44	0.888
E_LS	0.001	-0.003	0.008	0	1.546	1.553	0.888
V_MML	0.148	0.15	0.149	0.146	0.054	0.05	0.002
V_LS	0.277	0.276	0.279	0.277	0.117	0.106	0.004
RE (MML/LS)	54	54	53	53	46	47	53
n=100							
E_MML	-0.001	-0.002	0.005	0.001	1.353	1.356	0.897
E_LS	-0.003	-0.002	0.001	-0.005	1.565	1.566	0.897
V_MML	0.042	0.041	0.04	0.042	0.012	0.011	0
V_LS	0.083	0.082	0.074	0.078	0.039	0.035	0.001
RE (MML/LS)	51	50	54	53	30	33	43

Table D.7

ρ =0.2, Mixture Model (a), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.004	0.002	-0.012	-0.015	1.474	1.129	0.257
E_LS	0.006	0.002	-0.016	-0.015	1.513	1.042	0.283
V_MML	0.233	0.159	0.241	0.151	0.127	0.022	0.021
V_LS	0.4	0.176	0.416	0.168	0.198	0.021	0.027
RE (MML/LS)	58	91	58	90	64	104	78
n=30							
E_MML	-0.008	-0.006	0.006	-0.006	1.43	1.084	0.263
E_LS	-0.014	-0.007	0.009	-0.009	1.534	1.035	0.294
V_MML	0.151	0.104	0.149	0.103	0.074	0.014	0.013
V_LS	0.274	0.116	0.277	0.116	0.14	0.014	0.018
RE (MML/LS)	55	90	54	89	53	98	75
n=100							
E_MML	-0.013	-0.003	0	0.013	1.358	1.033	0.266
E_LS	-0.019	-0.003	0.001	0.011	1.565	1.031	0.307
V_MML	0.042	0.031	0.04	0.028	0.016	0.004	0.003
V_LS	0.081	0.034	0.083	0.033	0.045	0.004	0.005
RE (MML/LS)	52	89	49	87	35	85	65

Table D.8

ρ =0.9, Mixture Model (a), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.004	0.004	-0.012	-0.017	1.474	1.415	0.931
E_LS	0.006	0.006	-0.016	-0.02	1.513	1.435	0.941
V_MML	0.233	0.218	0.241	0.224	0.127	0.095	0.001
V_LS	0.4	0.354	0.416	0.366	0.198	0.15	0.001
RE (MML/LS)	58	62	58	61	64	64	98
n=30							
E_MML	-0.008	-0.009	0.006	0.002	1.43	1.371	0.935
E_LS	-0.014	-0.015	0.009	0.004	1.534	1.452	0.946
V_MML	0.151	0.141	0.149	0.14	0.074	0.056	0.001
V_LS	0.274	0.242	0.277	0.245	0.14	0.107	0.001
RE (MML/LS)	55	58	54	57	53	52	96
n=100							
E_MML	-0.013	-0.012	0	0.005	1.358	1.302	0.939
E_LS	-0.019	-0.017	0.001	0.005	1.565	1.477	0.953
V_MML	0.042	0.04	0.04	0.037	0.016	0.012	0
V_LS	0.081	0.071	0.083	0.071	0.045	0.034	0
RE (MML/LS)	52	56	49	52	35	36	89

Table D.9

ρ =0.2, Mixture Model (b), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.009	-0.002	-0.008	-0.013	1.051	1.543	0.143
E_LS	0.011	-0.003	-0.01	-0.017	0.987	1.536	0.137
V_MML	0.153	0.226	0.155	0.236	0.02	0.132	0.018
V_LS	0.165	0.384	0.167	0.412	0.02	0.197	0.024
RE (MML/LS)	93	59	93	57	102	67	75
n=30							
E_MML	0.003	-0.005	0.006	-0.011	1.036	1.462	0.145
E_LS	0.002	-0.005	0.005	-0.019	0.995	1.538	0.136
V_MML	0.1	0.148	0.099	0.143	0.013	0.073	0.012
V_LS	0.108	0.266	0.109	0.256	0.013	0.137	0.017
RE (MML/LS)	92	56	91	56	97	54	73
n=100							
E_MML	-0.009	-0.013	-0.006	0.005	1.008	1.356	0.147
E_LS	-0.011	-0.022	-0.005	0.005	0.996	1.551	0.127
V_MML	0.033	0.041	0.029	0.038	0.003	0.018	0.004
V_LS	0.036	0.08	0.031	0.074	0.004	0.05	0.005
RE (MML/LS)	91	52	94	51	87	35	70

Table D.10

ρ =0.9, Mixture Model (b), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.009	0.006	-0.008	-0.012	1.051	1.173	0.808
E_LS	0.011	0.007	-0.01	-0.016	0.987	1.128	0.791
V_MML	0.153	0.167	0.155	0.169	0.02	0.026	0.007
V_LS	0.165	0.208	0.167	0.214	0.02	0.033	0.01
RE (MML/LS)	93	80	93	79	102	79	65
n=30							
E_MML	0.003	0	0.006	0	1.036	1.138	0.82
E_LS	0.002	-0.001	0.005	-0.004	0.995	1.131	0.795
V_MML	0.1	0.108	0.099	0.109	0.013	0.015	0.004
V_LS	0.108	0.138	0.109	0.14	0.013	0.022	0.007
RE (MML/LS)	92	78	91	78	97	69	55
n=100							
E_MML	-0.009	-0.013	-0.006	-0.002	1.008	1.086	0.834
E_LS	-0.011	-0.019	-0.005	-0.002	0.996	1.129	0.793
V_MML	0.033	0.035	0.029	0.031	0.003	0.004	0.001
V_LS	0.036	0.045	0.031	0.04	0.004	0.007	0.002
RE (MML/LS)	91	77	94	78	87	54	40

Table D.11

ρ =0.2, Mixture Model (c), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.004	-0.017	0	-0.005	1.47	1.563	0.191
E_LS	0.008	-0.021	-0.002	-0.007	1.509	1.563	0.199
V_MML	0.227	0.246	0.237	0.247	0.125	0.132	0.019
V_LS	0.4	0.423	0.412	0.427	0.198	0.193	0.028
RE (MML/LS)	57	58	57	58	63	68	68
n=30							
E_MML	0.007	0.004	-0.011	-0.017	1.431	1.479	0.198
E_LS	0.013	-0.003	-0.013	-0.022	1.535	1.562	0.203
V_MML	0.145	0.149	0.146	0.154	0.073	0.075	0.012
V_LS	0.268	0.267	0.276	0.279	0.139	0.138	0.02
RE (MML/LS)	54	56	53	55	52	54	62
n=100							
E_MML	-0.002	-0.01	-0.007	-0.003	1.354	1.371	0.198
E_LS	0.004	-0.019	-0.015	0.003	1.557	1.579	0.2
V_MML	0.039	0.043	0.04	0.043	0.018	0.016	0.003
V_LS	0.08	0.081	0.081	0.087	0.048	0.044	0.006
RE (MML/LS)	48	53	50	50	37	36	53

Table D.12

ρ =0.9, Mixture Model (c), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.004	-0.004	0	-0.002	1.47	1.498	0.876
E_LS	0.008	-0.003	-0.002	-0.005	1.509	1.533	0.876
V_MML	0.227	0.232	0.237	0.237	0.125	0.095	0.005
V_LS	0.4	0.408	0.412	0.412	0.198	0.152	0.007
RE (MML/LS)	57	57	57	57	63	63	69
n=30							
E_MML	0.007	0.007	-0.011	-0.017	1.431	1.447	0.886
E_LS	0.013	0.009	-0.013	-0.02	1.535	1.55	0.884
V_MML	0.145	0.142	0.146	0.147	0.073	0.053	0.003
V_LS	0.268	0.261	0.276	0.274	0.139	0.103	0.005
RE (MML/LS)	54	55	53	54	52	52	60
n=100							
E_MML	-0.002	-0.006	-0.007	-0.007	1.354	1.359	0.895
E_LS	0.004	-0.005	-0.015	-0.011	1.557	1.566	0.893
V_MML	0.039	0.041	0.04	0.042	0.018	0.013	0.001
V_LS	0.08	0.082	0.081	0.085	0.048	0.035	0.001
RE (MML/LS)	48	50	50	50	37	37	49

Table D.13

ρ =0.2, Contamination Model (a), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	-0.004	-0.001	-0.005	0.001	0.988	1.108	0.18
E_LS	-0.005	-0.001	-0.008	0	0.941	1.012	0.187
V_MML	0.127	0.15	0.127	0.149	0.02	0.022	0.021
V_LS	0.147	0.16	0.148	0.16	0.02	0.02	0.024
RE (MML/LS)	87	93	86	93	100	110	88
n=30							
E_MML	0	-0.004	0.003	-0.003	0.967	1.065	0.181
E_LS	-0.001	-0.003	0.003	-0.004	0.942	1.006	0.187
V_MML	0.101	0.113	0.097	0.106	0.014	0.013	0.016
V_LS	0.101	0.113	0.097	0.106	0.014	0.013	0.016
RE (MML/LS)	85	92	86	92	96	101	88
n=100							
E_MML	-0.002	-0.007	0.002	-0.001	0.938	1.02	0.187
E_LS	-0.001	-0.006	0.002	-0.001	0.944	1.004	0.191
V_MML	0.024	0.03	0.024	0.03	0.004	0.004	0.004
V_LS	0.03	0.033	0.028	0.032	0.004	0.004	0.004
RE (MML/LS)	82	91	84	96	91	91	92

Table D.14

ρ =0.9, Contamination Model (a), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	-0.004	-0.004	-0.005	-0.003	0.988	1.016	0.874
E_LS	-0.005	-0.005	-0.008	-0.006	0.941	0.958	0.882
V_MML	0.127	0.128	0.127	0.132	0.02	0.018	0.002
V_LS	0.147	0.146	0.148	0.15	0.02	0.018	0.002
RE (MML/LS)	87	88	86	88	100	101	101
n=30							
E_MML	0	-0.002	0.003	0.001	0.967	0.988	0.878
E_LS	-0.001	-0.002	0.003	0	0.942	0.956	0.884
V_MML	0.086	0.09	0.083	0.086	0.013	0.011	0.001
V_LS	0.101	0.104	0.097	0.098	0.014	0.012	0.001
RE (MML/LS)	85	86	86	88	96	95	99
n=100							
E_MML	-0.002	-0.005	0.002	0.001	0.938	0.956	0.884
E_LS	-0.001	-0.004	0.002	0.001	0.944	0.957	0.888
V_MML	0.024	0.025	0.024	0.025	0.004	0.003	0
V_LS	0.03	0.029	0.028	0.028	0.004	0.004	0
RE (MML/LS)	82	85	84	88	91	91	97

Table D.15

ρ =0.2, Contamination Model (b), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.001	-0.005	-0.002	-0.004	1.051	1.046	0.201
E_LS	0.001	-0.006	-0.001	-0.007	0.988	0.967	0.204
V_MML	0.153	0.127	0.146	0.129	0.021	0.022	0.02
V_LS	0.163	0.146	0.158	0.147	0.02	0.021	0.023
RE (MML/LS)	93	87	93	88	103	107	86
n=30							
E_MML	-0.004	0.001	-0.01	0.002	1.032	1.004	0.206
E_LS	-0.002	0.002	-0.011	0	0.992	0.96	0.207
V_MML	0.104	0.086	0.097	0.085	0.013	0.013	0.014
V_LS	0.114	0.1	0.106	0.098	0.013	0.013	0.016
RE (MML/LS)	91	86	92	87	96	99	86
n=100							
E_MML	0.002	0	-0.006	-0.005	1.008	0.954	0.214
E_LS	0.005	0	-0.009	-0.006	0.995	0.954	0.212
V_MML	0.031	0.027	0.03	0.024	0.004	0.003	0.004
V_LS	0.033	0.032	0.032	0.029	0.004	0.004	0.005
RE (MML/LS)	91	84	93	83	92	91	86

Table D.16

ρ =0.9, Contamination Model (b), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0.001	-0.002	-0.002	-0.003	1.051	1.052	0.897
E_LS	0.001	-0.002	-0.001	-0.004	0.988	0.985	0.9
V_MML	0.153	0.148	0.146	0.141	0.021	0.018	0.001
V_LS	0.163	0.16	0.158	0.154	0.02	0.017	0.002
RE (MML/LS)	93	92	93	92	103	103	94
n=30							
E_MML	-0.004	-0.003	-0.01	-0.007	1.032	1.028	0.903
E_LS	-0.002	-0.001	-0.011	-0.009	0.992	0.987	0.903
V_MML	0.104	0.099	0.097	0.097	0.013	0.011	0.001
V_LS	0.114	0.109	0.106	0.105	0.013	0.012	0.001
RE (MML/LS)	91	91	92	92	96	95	90
n=100							
E_MML	0.002	0.002	-0.006	-0.007	1.008	0.998	0.909
E_LS	0.005	0.004	-0.009	-0.01	0.995	0.988	0.907
V_MML	0.031	0.029	0.03	0.029	0.004	0.003	0
V_LS	0.033	0.033	0.032	0.031	0.004	0.004	0
RE (MML/LS)	91	90	93	91	92	93	86

Table D.17

ρ =0.2, Contamination Model (c), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0	0.01	-0.008	-0.006	0.984	1.043	0.187
E_LS	-0.001	0.01	-0.01	-0.004	0.936	0.964	0.191
V_MML	0.126	0.128	0.133	0.127	0.02	0.022	0.02
V_LS	0.146	0.146	0.152	0.145	0.02	0.02	0.023
RE (MML/LS)	86	88	87	87	103	107	86
n=30							
E_MML	-0.002	-0.004	-0.005	0.002	0.965	1.001	0.192
E_LS	-0.005	-0.004	-0.004	0.001	0.94	0.957	0.195
V_MML	0.086	0.089	0.082	0.083	0.013	0.013	0.013
V_LS	0.102	0.103	0.097	0.096	0.013	0.013	0.015
RE (MML/LS)	85	86	85	87	96	99	85
n=100							
E_MML	-0.006	-0.01	0.003	-0.011	0.938	0.954	0.197
E_LS	-0.011	-0.01	0.003	-0.014	0.944	0.955	0.199
V_MML	0.024	0.025	0.025	0.026	0.004	0.004	0.004
V_LS	0.028	0.029	0.03	0.031	0.004	0.004	0.005
RE (MML/LS)	84	86	83	83	91	93	82

Table D.18

ρ =0.9, Contamination Model (c), b_x =1, b_y =1, b_u =1, b_v =1.							
n=20	μ_x	μ_y	μ_u	μ_v	σ_1	σ_2	ρ
E_MML	0	0.004	-0.008	-0.009	0.984	0.997	0.884
E_LS	-0.001	0.004	-0.01	-0.01	0.936	0.943	0.889
V_MML	0.126	0.125	0.133	0.13	0.02	0.017	0.002
V_LS	0.146	0.144	0.152	0.149	0.02	0.017	0.002
RE (MML/LS)	86	87	87	88	103	102	98
n=30							
E_MML	-0.002	-0.004	-0.005	-0.003	0.965	0.973	0.89
E_LS	-0.005	-0.006	-0.004	-0.002	0.94	0.944	0.894
V_MML	0.086	0.086	0.082	0.082	0.013	0.011	0.001
V_LS	0.102	0.101	0.097	0.097	0.013	0.011	0.001
RE (MML/LS)	85	85	85	85	96	95	94
n=100							
E_MML	-0.006	-0.009	0.003	-0.003	0.938	0.942	0.896
E_LS	-0.011	-0.013	0.003	-0.004	0.944	0.947	0.897
V_MML	0.024	0.024	0.025	0.026	0.004	0.003	0
V_LS	0.028	0.028	0.03	0.032	0.004	0.003	0
RE (MML/LS)	84	84	83	82	91	88	92

APPENDIX E

FORTRAN PROGRAMS

We will only show part of the program for Student's t distribution. The one for the Generalized Logistic is similar and will not be displayed here.

Note: Sigma in the program is $\sigma_{2.1}$

E.1 Student's t

```
c    ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
c            A PROCEDURE THAT CALCULATES THE LS ESTIMATORS
c      takes the sample and df as input, produces all LSE
c    ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
      subroutine LS(n,rx,ru,ry,rv,x,y,u,v,c1_w, c2_w, mu_xH_LS1,
     & mu_yH_LS1,mu_uH_LS1,mu_vH_LS1,mu_ygxH_LS1, mu_vguH_LS1,
     & sigma1H_LS1,sigma2H_LS1,sigmaH_LS1,rhoH_LS1,thetaH_LS1)

       real x(300),y(300),u(300),v(300)
       real wls1(300),wls2(300)
       real c1_w(n),c2_w(n)
       real mu_xH_LS1,mu_yH_LS1,mu_uH_LS1, mu_vH_LS1
      real mu_ygxH_LS1, mu_vguH_LS1,sigma1H_LS1,sigma2H_LS1,sigmaH_LS1
       real rhoH_LS1,thetaH_LS1
       real xbar,ubar,vbar,ybar,xd_bar,ud_bar,vd_bar,yd_bar
       real mean_wls1,mean_wls2

       xbar = 0.0
       ubar = 0.0
       ybar = 0.0
       vbar = 0.0

      do 1068 i=1,n
       c1_w(i) = 1.0
       c2_w(i) = 1.0
       xbar = xbar + x(i)
```

```
      ybar = ybar + y(i)
      ubar = ubar + u(i)
      vbar = vbar + v(i)
1068    continue

      xbar = xbar/(1.0*n)
      ybar = ybar/(1.0*n)
      ubar = ubar/(1.0*n)
      vbar = vbar/(1.0*n)

      yd_bar = 0.0
      xd_bar =0.0
      ud_bar = 0.0
      vd_bar = 0.0
      c1_w_sum = 0.0
      c2_w_sum = 0.0

      do  2901 i=1,n
      xd_bar = xd_bar + c1_w(i)*x(i)
      yd_bar = yd_bar + c1_w(i)*y(i)
      ud_bar = ud_bar + c2_w(i)*u(i)
      vd_bar = vd_bar + c2_w(i)*v(i)
      c1_w_sum = c1_w_sum + c1_w(i)
      c2_w_sum = c2_w_sum + c2_w(i)
2901  continue

      yd_bar = yd_bar/c1_w_sum
      xd_bar = xd_bar/c1_w_sum
      vd_bar = vd_bar/c2_w_sum
      ud_bar = ud_bar/c2_w_sum

c     ******** SAMPLE VARIANCES AND COVARIANCES *******************
      sx2 = 0.0
      sy2 = 0.0
      su2 = 0.0
      sv2 = 0.0
c     sum of (xi-xbar)(yi-ybar)
      sxy = 0.0
      suv = 0.0
c     *the two parts of the denominator of thetaH_LS1:
      td1 = 0.0
      td2 = 0.0

      do 1069 i=1,n
      sx2 = sx2 + ((x(i) - xbar)**2)
      sy2 = sy2 + ((y(i) - ybar)**2)
      su2 = su2 + ((u(i) - ubar)**2)
      sv2 = sv2 + ((v(i) - vbar)**2)
      sxy = sxy + c1_w(i)*((x(i)-xd_bar)*(y(i)-yd_bar))
      suv = suv + c2_w(i)*((u(i)-ud_bar)*(v(i)-vd_bar))
      td1 = td1 + c1_w(i)* ((x(i) - xd_bar)**2)
```

```
      td2 = td2 + c2_w(i)* ((u(i) - ud_bar)**2)
1069  continue

      sx2 = sx2/(1.0*n-1.0)
      sy2 = sy2/(1.0*n-1.0)
      su2 = su2/(1.0*n-1.0)
      sv2 = sv2/(1.0*n-1.0)
      sxy = sxy/(1.0*n-1.0)
      suv = suv/(1.0*n-1.0)
      td1 = td1/(1.0*n-1.0)
      td2 = td2/(1.0*n-1.0)

      thetaH_LS1 = (sxy+suv)/(td1+td2)

      sigma1H_LS1 = sqrt((sx2+su2)/(rx/(rx-2.0)+ru/(ru-2.0)))

c     ** Finding sigmaH_LS1 -- only for the case where c1i=c2i=1
c     ** we do this by finding wls1i = yi - theta*xi,
c     ** and wls2i= vi - theta*ui.
c     ** Now, note that wls1bar is ybar-theta*xbar
c     ** Thus in finding var(wls) we can find the numerator of sigmaH_LS1

      mean_wls1 = 0.0
      mean_wls2 = 0.0

     do 200 i = 1,n
      wls1(i) = y(i) - thetaH_LS1*x(i)
      wls2(i) = v(i) - thetaH_LS1*u(i)
      mean_wls1 = mean_wls1 + wls1(i)
      mean_wls2 = mean_wls2 + wls2(i)
200   continue

c     means of wls's
      mean_wls1 = mean_wls1/(1.0*n)
      mean_wls2 = mean_wls2/(1.0*n)

      var_wls1 = 0.0
      var_wls2 = 0.0

     do 201 i=1,n
      var_wls1 = var_wls1 + ((wls1(i)-mean_wls1)**2)
      var_wls2 = var_wls2 + ((wls2(i)-mean_wls2)**2)
201   continue

c     variances of wls's adjusted for bias
      var_wls1 = var_wls1/(1.0*n-2.0)
      var_wls2 = var_wls2/(1.0*n-2.0)

      sigmaH_LS1 = sqrt((var_wls1+var_wls2)/
     &((ry/(ry-2))+(rv/(rv-2))))
```

```
      sigma2H_LS1 =  sqrt((sigmaH_LS1**2)+
     &(thetaH_LS1**2)*(sigma1H_LS1**2))

      rhoH_LS1 = thetaH_LS1*sigma1H_LS1/sigma2H_LS1

      mu_xH_LS1 = xbar
      mu_uH_LS1 = ubar

      mu_yH_LS1 = yd_bar - thetaH_LS1*(xd_bar-mu_xH_LS1)

      mu_vH_LS1 = vd_bar - thetaH_LS1*(ud_bar-mu_uH_LS1)

      mu_ygxH_LS1 = yd_bar - thetaH_LS1*xd_bar
      mu_vguH_LS1 = vd_bar - thetaH_LS1*ud_bar
   return
   end

c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
c           CALCULATES THE MML ESTIMATORS
c      takes the sample and df as input, produces all estimates
c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
     subroutine MML(n,rx,ru,ry,rv,thetaH_LS1,
     &x,y,u,v,c1_w,c2_w,mu_xH_MML1,mu_yH_MML1,mu_uH_MML1,
     &mu_vH_MML1,mu_ygxH_MML1,mu_vguH_MML1,
     &sigma1H_MML1,sigma2H_MML1,sigmaH_MML1,rhoH_MML1,thetaH_MML1,
     &m11, m12,m21,m22,alpha1,alpha2,beta1,beta2,gamma1,gamma2,
     &delta1,delta2,alpha_w,beta_w,delta_w,gamma_w,t1,t2,s1,s2)

      real x(300),y(300),u(300),v(300),xo(300),uo(300)
      real z1(300),z2(300),w1(300),w2(300)
      real xc(300),yc(300),uc(300),vc(300)
      real wls1(300),wls2(300)
      real c1_w(300),c2_w(300)
      real mu_xH_MML1,mu_yH_MML1,mu_uH_MML1, mu_vH_MML1
      real mu_ygxH_MML1, mu_vguH_MML1,sigma1H_MML1,sigma2H_MML1
      real sigmaH_MML1,rhoH_MML1,thetaH_MML1,qi
      real t1(300),t2(300),s1(300),s2(300)
      real beta1(300),beta2(300),alpha1(300),alpha2(300)
      real gamma1(300),gamma2(300),delta1(300),delta2(300)
      real beta_w(300),alpha_w(300),gamma_w(300),delta_w(300)
      real m11, m12,m21,m22,k11,k21
      real K,K1,K2,D,D1,D2,S,SS1,SS2

c     **** Sorting the X's and U's *****

      CALL SVRGN (n, x, xo)
      CALL SVRGN (n, u, uo)


c     *** Finding Beta's alpha's, gamma's, delta's and s's and t's ****
c     *Initializing:
```

```
      do 1006 i=1,n
      qi = (1.0*i)/(1.0*n+1.0)
      t1(i) = tin(qi,rx)
      beta1(i) = (1.0-((t1(i)**2)/rx))/((1.0+((t1(i)**2)/rx))**2)
      alpha1(i) = (2.0*(t1(i)**3)/rx)/((1.0+((t1(i)**2)/rx))**2)

      if (beta1(i).lt.0) then
      beta1(i) = 1.0/((1.0+((t1(i)**2)/rx))**2)
      alpha1(i) = ((t1(i)**3)/rx)/((1.0+((t1(i)**2)/rx))**2)
      endif

      t2(i) = tin(qi,ry)
      beta2(i) = (1.0-((t2(i)**2)/ry))/((1.0+((t2(i)**2)/ry))**2)
      alpha2(i) = (2.0*(t2(i)**3)/ry)/((1.0+((t2(i)**2)/ry))**2)

      if (beta2(i).lt.0) then
      beta2(i) = 1.0/((1.0+((t2(i)**2)/ry))**2)
      alpha2(i) = ((t2(i)**3)/ry)/((1.0+((t2(i)**2)/ry))**2)
      endif

      s1(i) = tin(qi,ru)
      gamma1(i) = (1.0-((s1(i)**2)/ru))/((1.0+((s1(i)**2)/ru))**2)
      delta1(i) = (2.0*(s1(i)**3)/ru)/((1.0+((s1(i)**2)/ru))**2)

      if (gamma1(i).lt.0) then
      gamma1(i) = 1.0/((1.0+((s1(i)**2)/ru))**2)
      delta1(i) = ((s1(i)**3)/ru)/((1.0+((s1(i)**2)/ru))**2)
      endif

      s2(i) = tin(qi,rv)
      gamma2(i) = (1.0-((s2(i)**2)/rv))/((1.0+((s2(i)**2)/rv))**2)
      delta2(i) = (2.0*(s2(i)**3)/rv)/((1.0+((s2(i)**2)/rv))**2)

      if (gamma2(i).lt.0) then
      gamma2(i) = 1.0/((1.0+((s2(i)**2)/rv))**2)
      delta2(i) = ((s2(i)**3)/rv)/((1.0+((s2(i)**2)/rv))**2)
      endif

      beta_w(i) = c1_w(i)*beta2(i)
      alpha_w(i) = sqrt(c1_w(i))*alpha2(i)
      gamma_w(i) = c2_w(i)*gamma2(i)
      delta_w(i) = sqrt(c2_w(i))*delta2(i)

1006  continue

c     sum of beta1i's
      m11 = 0.0
c     sum of beta2i's
      m12 = 0.0
c     sum of gamma1i's
```

```
      m21 = 0.0
c     sum of gamma2i's
      m22 = 0.0
c     sum of beta1i*xoi
      k11 = 0.0
c     sum of gamma1i*uoi
      k21 = 0.0

      do 3011 i=1,n
      m11 = m11 + beta1(i)
      m12 = m12 + beta2(i)
c     sum of beta1i*xo(i) will divide by m11 later
      k11 = k11 + beta1(i)*xo(i)
      m21 = m21 + gamma1(i)
      m22 = m22 + gamma2(i)
c     sum of gamma1i*u(i)/m21 will divide by m21 later
      k21 = k21 + gamma1(i)*uo(i)

3011  continue

c     $$$$$$$$$$$$$$$$$ 1. mu_xH and mu_uH  $$$$$$$$$$$$$$$$$$$$$$$$$

      mu_xH_MML1 = k11/m11
      mu_uH_MML1 = k21/m21

c     $$$$$$$$$$$$$$ 2. SIGMA1H_MML1 $$$$$$$$$$$$$$$$$$$$$$$$$$$$$$

c     to calculate sigma1H_MML1
      b1a = 0.0
      b1b = 0.0
c     ** the first and second parts of C1 in sigma1H_MML1
      C1a = 0.0
      C1b = 0.0
      B1 = 0.0
      C1 = 0.0

      do 1177 i =1,n
      b1a = b1a + alpha1(i)*xo(i)
      b1b = b1b + delta1(i)*uo(i)
      C1a = C1a + beta1(i)*((xo(i)-mu_xH_MML1)**2)
      C1b = C1b + gamma1(i)*((uo(i)-mu_uH_MML1)**2)
1177  continue

c     *the first and second parts of B1 in sigma1H_MML1:
      b1a = ((rx+1.0)/(1.0*rx))*b1a
      b1b = ((ru+1.0)/(1.0*ru))*b1b
c     *the first and second parts of C1 in sigma1H_MML1:
      C1a = ((rx+1.0)/(1.0*rx))*C1a
      C1b = ((ru+1.0)/(1.0*ru))*C1b
      B1 = b1a + b1b
      C1 = C1a + C1b
```

```
      sigma1H_MML1=(B1+sqrt((B1**2)+(8.0*n*C1)))/
      &(4.0*sqrt(1.0*n*(n-2.0)))


c     The errors so we can order according to them (no need for mu's-constants)
      test = thetaH_LS1
      call concom(mu_x,mu_y,mu_u,mu_v,sigma1,sigma,
      &theta,rho,mu_ygx,mu_vgu,rx,ry,ru,rv,x,y,u,v,xc,uc,yc,vc,
      &z1,z2,w1,w2,n,test)

c     *** Calculating x[.]bar, u[.]bar ...etc
      xdot_bar = 0.0
      ydot_bar = 0.0
      udot_bar = 0.0
      vdot_bar = 0.0

      do 1184 i = 1,n
      xdot_bar = xdot_bar + beta_w(i)*xc(i)
      ydot_bar = ydot_bar + beta_w(i)*yc(i)
      udot_bar = udot_bar + gamma_w(i)*uc(i)
      vdot_bar = vdot_bar + gamma_w(i)*vc(i)

1184  continue

      xdot_bar = xdot_bar/m12
      ydot_bar = ydot_bar/m12
      udot_bar = udot_bar/m22
      vdot_bar = vdot_bar/m22


c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
c     CALCULATING SIGMA_HAT and THETA_HAT
c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
      K = 0.0
c     first part of K
      K1 = 0.0
c     second part of K
      K2 = 0.0
      D = 0.0
c     first and second parts of D
      D1 = 0.0
      D2 = 0.0
      S = 0.0
c     first and second parts of S
      SS1 = 0.0
      SS2 = 0.0

      do 1186 i = 1,n
      SS1 = SS1 + ((ry+1.0)/(1.0*ry))*beta_w(i)*((xc(i) - xdot_bar)**2)
```

```
      SS2 = SS2 + ((rv+1.0)/rv)* gamma_w(i)*((uc(i) - udot_bar)**2)
      D1 = D1 + ((ry+1.0)/(1.0*ry))*alpha_w(i)*(xc(i) - xdot_bar)
      D2 = D2 + ((rv+1.0)/(1.0*rv))*delta_w(i)*(uc(i) - udot_bar)
      K1 = K1 + ((ry+1.0)/(1.0*ry))*beta_w(i)*(xc(i) - xdot_bar)*yc(i)
      K2 = K2 + ((rv+1.0)/(1.0*rv))*gamma_w(i)*(uc(i) - udot_bar)*vc(i)

1186  continue

c     The S, D, and K that are in the thetaH_MML1
      S = SS1 + SS2
      D = (D1 + D2)/S
      K = (K1 + K2)/S

c     The things needed for sigmaH: (both parts of B2 and C2
      B2a = 0.0
      B2b = 0.0
      C2a = 0.0
      C2b = 0.0

      do 1187 i = 1,n
      B2a=B2a+((ry+1)/ry)*alpha_w(i)*(yc(i)- ydot_bar -
      &(K*(xc(i)- xdot_bar)))
      B2b=B2b+((rv+1)/rv)*delta_w(i)*(vc(i)- vdot_bar -
      &(K*(uc(i)- udot_bar)))
      C2a=C2a+((ry+1)/ry)*beta_w(i)*((yc(i)-ydot_bar-
      &(K*(xc(i)-xdot_bar)))**2)
      C2b=C2b+((rv+1)/rv)*gamma_w(i)*((vc(i)-vdot_bar-
      &(K*(uc(i)-udot_bar)))**2)
1187  continue

c     B and C that are in sigmaH_MML1
      B2 = B2a + B2b
      C2 = C2a + C2b

      sigmaH_MML1 = (B2 + sqrt((B2**2)+(8.0*n*C2)))/
      &(4.0*sqrt(1.0*n*(n-4.0)))

      thetaH_MML1 = K + D*sigmaH_MML1

c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
c      CALCULATING mu_ygxH_MML1, mu_vguH_MML1
c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

      alpha_sum = 0.0
      delta_sum = 0.0

      do 5000 i=1,n
      alpha_sum = alpha_sum + alpha_w(i)
      delta_sum = delta_sum + delta_w(i)
5000  continue
```

```
c     **if we assume c1i's are one, the sum of
c       alphas and deltas are both zero

      mu_ygxH_MML1 = ydot_bar - (thetaH_MML1*xdot_bar)
      mu_vguH_MML1 = vdot_bar - (thetaH_MML1*udot_bar)

c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
c      CALCULATING mu_yH_MML1, mu_vH_MML1, sigma2H and rhoH
c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

      mu_yH_MML1 = ydot_bar - thetaH_MML1*(xdot_bar - mu_xH_MML1)

      mu_vH_MML1 = vdot_bar - thetaH_MML1*(udot_bar - mu_uH_MML1)

      sigma2H_MML1 = sqrt((sigmaH_MML1**2)+(thetaH_MML1**2)*
     &(sigma1H_MML1**2))

      rhoH_MML1 = thetaH_MML1*sigma1H_MML1/sigma2H_MML1

      return
      end

c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
c          PERFORMES A HYPOTHESIS TEST AND FINDS POWER
c           compares That to Ttelda
c           for testing equality of two mean vectors
c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
      subroutine hypoth_eqmeans(n,iter,rx,ry,ru,rv,mu_x,mu_u,
     &mu_y,mu_v,mu_ygx,mu_vgu,sigma1,sigma2,sigma,rho,theta,
     &c1_w,c2_w,I1inv,power_MML,power_LS)

      real alpha1(300),beta1(300),alpha2(300),beta2(300)
      real gamma1(300),gamma2(300),delta1(300),delta2(300)
      real alpha_w(300),beta_w(300),gamma_w(300),delta_w(300)
      real t1(300),t2(300),s1(300),s2(300)
      real I1inv(7,7)
      real TS_MML(iter), TS_LS(iter)
      real power_MML, power_LS
      real c1_w(300),c2_w(300)
      real sigma_xuH,sigma_yvH,sigma_xuT,sigma_yvT
      real pval_MML, al, d1,d2
      real CV_MML, CV_LS
      real temp, temp1, temp2, chiin
      real x(300),y(300),u(300),v(300),xo(300),uo(300)
      real z1(300),z2(300),w1(300),w2(300)
      real mu_xH_MML1,mu_yH_MML1,mu_uH_MML1,mu_vH_MML1
      real mu_ygxH_MML1,mu_vguH_MML1,sigma1H_MML1,sigma2H_MML1
      real sigmaH_MML1,rhoH_MML1,thetaH_MML1
      real mu_xH_LS1,mu_yH_LS1,mu_uH_LS1,mu_vH_LS1,mu_ygxH_LS1
      real mu_vguH_LS1,sigma1H_LS1,sigma2H_LS1,sigmaH_LS1
      real rhoH_LS1,thetaH_LS1,thetaH_LS
```

```
      real v11, v12,v21,v22, per
      real m11,m12,m21,m22
      real c1bar, c2bar

      c1bar = 0.0
      c2bar = 0.0
      do 1666 i=1,n
      c1_w(i) = 1.0
      c2_w(i) = 1.0
      c1bar = c1bar + c1_w(i)
      c2bar = c2bar + c2_w(i)
1666 continue

      c1bar = c1bar/(1.0*n)
      c2bar = c2bar/(1.0*n)

      DF = 2.0
c     If we just want to use the hypothetical chi-square distribution
c     CV_MML = chiin(0.95,DF)
c     CV_LS = chiin(0.95,DF)

c      If we want to use a simulated critical value to make sure the
c      type I error is 0.05 we use:
c     call findCV for usual sample with no deviations
c     call findCV3 for contamination3 model
c     findCV2 for the outlier3 model and,
c     findCV1 for the mixture3 model

      call findCV(n,CV_MML,CV_LS)

      write(8,*) '  d       MML        LS'
      write(8,*) '--------------------------------'

      do 3456 m = 1,20
      d1 = m*0.1
      d1 = d1 - 0.1

      power_MML = 0.0
      power_LS = 0.0

      do 188 i=1,iter

c     **generate numbers assuming H1 is true.  1. usual sample
      call sample(mu_x,mu_y,mu_u,mu_v,sigma1,sigma,
      &theta,rho,mu_ygx,mu_vgu,rx,ry,ru,rv,x,y,u,v,z1,z2,w1,w2,n)

c     ** 2. Change the x's and Y/X to match the alternative hypothesis:
      do 3649 j=1,n
      x(j) = x(j) + d1
      z2(j) = z2(j)+(d1/sigma)
      y(j)=mu_ygx+ theta*x(j) + sigma*z2(j)
```

```
3649  continue

      CALL LS(n,rx,ru,ry,rv,x,y,u,v,c1_w,c2_w, mu_xH_LS1,
     & mu_yH_LS1,mu_uH_LS1,mu_vH_LS1,mu_ygxH_LS1,mu_vguH_LS1,
      &sigma1H_LS1,sigma2H_LS1,sigmaH_LS1,rhoH_LS1,thetaH_LS1)

c     calculating the MMLE in two iterations:
c     -------first iteration using thetaH_LS1 to order------------------------------------
      thetaH_LS = thetaH_LS1
      CALL MML(n,rx,ru,ry,rv,thetaH_LS,
      &x,y,u,v,c1_w,c2_w,mu_xH_MML1,mu_yH_MML1,mu_uH_MML1,
      &mu_vH_MML1,mu_ygxH_MML1,mu_vguH_MML1,
      &sigma1H_MML1,sigma2H_MML1,sigmaH_MML1,rhoH_MML1,thetaH_MML1,
      &m11, m12,m21,m22,alpha1,alpha2,beta1,beta2,gamma1,gamma2,
      &delta1,delta2,alpha_w,beta_w,delta_w,gamma_w,t1,t2,s1,s2)

c     -------second iteration using thetaH_MML1 to order------------------------------------
      thetaH_LS = thetaH_MML1
      CALL MML(n,rx,ru,ry,rv,thetaH_LS,
      &x,y,u,v,c1_w,c2_w,mu_xH_MML1,mu_yH_MML1,mu_uH_MML1,
      &mu_vH_MML1,mu_ygxH_MML1,mu_vguH_MML1,
      &sigma1H_MML1,sigma2H_MML1,sigmaH_MML1,rhoH_MML1,thetaH_MML1,
      &m11, m12,m21,m22,alpha1,alpha2,beta1,beta2,gamma1,gamma2,
      &delta1,delta2,alpha_w,beta_w,delta_w,gamma_w,t1,t2,s1,s2)

c     ** calc the test statistics
      sigma_xuH = (sigma1H_MML1**2)*( (rx/((rx+1.0)*m11)) +
      &(ru/((ru+1.0)*m21)))

      sigma_yvH = (sigmaH_MML1**2)*( (ry/((ry+1.0)*m12)) +
      &(rv/((rv+1.0)*m22)))

c         ***T_hat2:
       TS_MML(i) =(((mu_xH_MML1-mu_uH_MML1)**2)/sigma_xuH)+
      &(((mu_ygxH_MML1 - mu_vguH_MML1)**2)/sigma_yvH)

      sigma_xuT = (sigma1H_LS1**2)*rx/(n*(rx-2.0)) +
      &(sigma1H_LS1**2)*ru/(n*(ru-2.0))

      sigma_yvT = (sigmaH_LS1**2)*ry/(c1bar*n*(ry-2.0)) +
      &(sigmaH_LS1**2)*rv/(c2bar*n*(rv-2.0))

c         ***T_telda2:
      TS_LS(i) = (((mu_xH_LS1 - mu_uH_LS1)**2)/sigma_xuT) +
      &(((mu_ygxH_LS1 - mu_vguH_LS1)**2)/sigma_yvT)

c     ** if we do reject H0:
      if (TS_MML(i).gt.CV_MML) then
      power_MML = power_MML + 1.0
      endif
```

```
      if (TS_LS(i).gt.CV_LS) then
      power_LS = power_LS + 1.0
      endif

188   continue

      power_MML = power_MML/(1.0*iter)
      power_LS = power_LS/(1.0*iter)

      write(8,432) d1, power_MML, power_LS
432   format(f8.3,f8.3,f8.3)

3456  continue

      return
      end

c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
c                    CREATS THE SAMPLES
c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

c     ***** 1.  sample from t-distribution: *****************

      subroutine sample(mu_x,mu_y,mu_u,mu_v,sigma1,sigma,
     &theta,rho,mu_ygx,mu_vgu,rx,ry,ru,rv,x,y,u,v,z1,z2,w1,w2,n)

      real z1(300),z2(300),w1(300),w2(300)
      real x(300),y(300),u(300),v(300)
      real mu_x,mu_y,mu_u,mu_v,sigma1,sigma
      real theta,rho,mu_ygx,mu_vgu

c     generates random numbers from T-dist:

      CALL RNSTT (n, ru, w1)
      CALL RNSTT (n, rv, w2)
      CALL RNSTT (n, rx, z1)
      CALL RNSTT (n, ry, z2)

      do 1000 i=1,n
      x(i)=mu_x+sigma1*z1(i)
      u(i)=mu_u+sigma1*w1(i)
      y(i)=mu_ygx+theta*x(i)+sigma*z2(i)
      v(i)=mu_vgu+theta*u(i)+sigma*w2(i)
1000  continue

      return
      end

c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
c     ** 3c. outlier model (changing y's and v's)
c       (n-r)t(rx) sigma1 and (r)t(rx) 4sigma1
```

```
c       (n-r)t(ru) sigma1 and (r)t(ru) 4sigma1
c       (n-r)t(ry) sigma and (r)t(ry)4sigma
c       (n-r)t(rv) sigma and (r)t(rv)4sigma
c     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
      subroutine outlier3(mu_x,mu_y,mu_u,mu_v,sigma1,sigma,
     &theta,rho,mu_ygx,mu_vgu,rx,ry,ru,rv,x,y,u,v,z1,z2,w1,w2,n)

      real z1(300),z2(300),w1(300),w2(300)
      real x(300),y(300),u(300),v(300)
      real mu_x,mu_y,mu_u,mu_v,sigma1,sigma
      real theta,rho,mu_ygx,mu_vgu

c     generates random numbers from T-dist:
      CALL RNSTT (n, ru, w1)
      CALL RNSTT (n, rv, w2)
      CALL RNSTT (n, rx, z1)
      CALL RNSTT (n, ry, z2)

      r = int(0.1*n+0.5)

c     find x's and u's:
      do 101 i=1,n
      x(i)=mu_x+sigma1*z1(i)
      u(i)=mu_u+sigma1*w1(i)
101   continue

c     ** Change some of the x, and u values to be outliers
      do 113 i = 1,r
      x(i)=mu_x + 4.0*sigma1*z1(i)
      u(i)=mu_u + 4.0*sigma1*w1(i)
113   continue

c     find y's and v's
      do 103 i=1,n
      y(i)=mu_ygx + theta*x(i) + sigma*z2(i)
      v(i)=mu_vgu + theta*u(i) + sigma*w2(i)
103   continue

c     change some of the y's,v's to be outliers
      do 102 i = 1,r
      y(i)=mu_ygx+theta*x(i)+4.0*sigma*z2(i)
      v(i)=mu_vgu+theta*u(i)+4.0*sigma*w2(i)
102   continue

      return
      end
```

APPENDIX F

INFORMATION MATRIX (UNEQUAL VAR-COV MATRICES)

Here we give the information matrix in the case where the covariance matrices for both populations are not the same.

F.1 Generalized Logistic

The elements of the information matrix $I_1(\mu_x, \sigma_x, \mu_{y/x}, \sigma_{2.11},\ \theta_{yx})$ are as follows (see Sazak et al. (2006)):

1. $I_{\mu_x\mu_x} = \dfrac{n}{\sigma_x^2}\left[\dfrac{b_x}{b_x+2}\right].$

$$I_{\mu_x\sigma_x} = \frac{n}{\sigma_x^2}\left[\frac{b_x}{(b_x+2)}\big(\psi(b_x+1)-\psi(2)\big)\right],$$

$$I_{\mu_x\mu_{y/x}} = I_{\mu_x\sigma_{2.11}} = I_{\mu_x\theta_{yx}} = 0.$$

2. $I_{\sigma_x\sigma_x} = \dfrac{n}{\sigma_x^2}\left[1+\dfrac{b_x}{b_x+2}\Big(\psi'(b_x+1)+\psi'(2)+\big(\psi(b_x+1)-\psi(2)\big)^2\Big)\right],$

$$I_{\sigma_x\mu_{y/x}} = I_{\sigma_x\sigma_{2.11}} = I_{\sigma_x\theta_{yx}} = 0.$$

3. $I_{\mu_{y/x}\mu_{y/x}} = \dfrac{n}{\sigma_{2.11}^2}\dfrac{b_y}{(b_y+2)},$

$$I_{\mu_{y/x}\sigma_{2.11}} = \frac{n}{\sigma_{2.11}^2}\frac{b_y}{(b_y+2)}\big(\psi(b_y+1)-\psi(2)\big),$$

$$I_{\mu_{y/x}\theta_{yx}} = \frac{n\sigma_1}{\sigma_{2.11}^2}\frac{b_y}{(b_y+2)}(\psi(b_x)-\psi(1)).$$

4. $I_{\sigma_{2.11}\sigma_{2.11}} = \frac{n}{\sigma_{2.11}^2}\left\{1 + \frac{b_y}{b_y+2}\left(\psi'(b_y+1) + \psi'(2) + \left(\psi(b_y+1) - \psi(2)\right)^2\right)\right\}$,

$$I_{\sigma_{2.11}\theta_{yx}} = \frac{n\sigma_x}{\sigma_{2.11}^2}\left\{\frac{b_y}{b_y+2}\left[\left(\psi(b_y+1) - \psi(2)\right)\left(\psi(b_x) - \psi(1)\right)\right]\right\}.$$

5. $I_{\theta_{yx}\theta_{yx}} = \frac{n\sigma_x^2}{\sigma_{2.1`}^2}\left\{\frac{b_y}{b_y+2}\left[\psi'(b_x) + \psi'(1) + \left(\psi(b_x) - \psi(1)\right)^2\right]\right\}$.

The elements of the Fisher information matrix $I_1(\mu_x,\ \sigma_x, \mu_y,\ \sigma_y,\ \rho_{yx})$ for estimating μ_x, σ_x, μ_y, σ_y, ρ_{yx} are given by:

1. $I_{\mu_x\mu_x} = \frac{n}{\sigma_x^2}\left[\frac{b_x}{b_x+2} + \frac{\rho_{yx}^2}{1-\rho_{yx}^2}\frac{b_y}{(b_y+2)}\right]$,

$$I_{\mu_x\sigma_x} = \frac{n}{\sigma_x^2}\left[\frac{b_x}{b_x+2}\left(\psi(b_x+1) - \psi(2)\right) + \frac{\rho_{yx}^2}{1-\rho_{yx}^2}\frac{b_y}{(b_y+2)}\left(\psi(b_x) - \psi(1)\right)\right],$$

$$I_{\mu_x\mu_y} = \frac{-n\rho_{yx}}{\sigma_x\sigma_y(1-\rho_{yx}^2)}\frac{b_y}{(b_y+2)},$$

$$I_{\mu_x\sigma_y} = \frac{-n\rho_{yx}}{\sigma_x\sigma_y\sqrt{1-\rho_{yx}^2}}\frac{b_y}{(b_y+2)}\left[\left(\psi(b_y+1) - \psi(2)\right) + \frac{\rho_{yx}}{\sqrt{1-\rho_{yx}^2}}\left(\psi(b_x) - \psi(1)\right)\right],$$

$$I_{\mu_x\rho_{yx}} = \frac{-n\rho_{yx}}{\sigma_x(1-\rho_{yx}^2)}\frac{b_y}{(b_y+2)}\left[\left(\psi(b_x) - \psi(1)\right) - \frac{\rho_{yx}}{\sqrt{1-\rho_{yx}^2}}\left(\psi(b_y+1) - \psi(2)\right)\right].$$

2. $I_{\sigma_x\sigma_x} = \frac{n}{\sigma_x^2}\left[1 + \frac{b_x}{b_x+2}\left(\psi'(b_x+1) + \psi'(2) + \left(\psi(b_x+1) - \psi(2)\right)^2\right)\right.$

$$\left. + \frac{\rho_{yx}^2}{1-\rho_{yx}^2}\frac{b_y}{(b_y+2)}\left(\psi'(b_x) + \psi'(1) + \left(\psi(b_x) - \psi(1)\right)^2\right)\right],$$

$$I_{\sigma_x\mu_y} = \frac{-n\rho_{yx}}{\sigma_x\sigma_y(1-\rho_{yx}^2)}\frac{b_y}{(b_y+2)}\left(\psi(b_x) - \psi(1)\right),$$

$$I_{\sigma_x\sigma_y} = \frac{-n\rho_{yx}}{\sigma_x\sigma_y\sqrt{1-\rho_{yx}^2}}\left\{\frac{b_y}{(b_y+2)}\left[\left(\psi(b_x)-\psi(1)\right)\left(\psi(b_y+1)-\psi(2)\right)\right.\right.$$

$$\left.\left.+\frac{\rho_{yx}}{\sqrt{1-\rho_{yx}^2}}\left(\psi'(b_x)+\psi'(1)+\left(\psi(b_x)-\psi(1)\right)^2\right)\right]\right\},$$

$$I_{\sigma_x\rho_{yx}} = \frac{-n\rho_{yx}}{\sigma_x(1-\rho_{yx}^2)}\left\{\frac{b_y}{(b_y+2)}\left[\left(\psi'(b_x)+\psi'(1)+\left(\psi(b_x)-\psi(1)\right)^2\right)\right.\right.$$

$$\left.\left.-\frac{\rho_{yx}}{\sqrt{1-\rho_{yx}^2}}\left(\psi(b_x)-\psi(1)\right)\left(\psi(b_y+1)-\psi(2)\right)\right]\right\}.$$

3. $$I_{\mu_y\mu_y} = \frac{n}{\sigma_y^2(1-\rho_{yx}^2)}\frac{b_y}{(b_y+2)},$$

$$I_{\mu_y\sigma_y} = \frac{n}{\sigma_y^2\sqrt{1-\rho_{yx}^2}}\frac{b_y}{(b_y+2)}\left[\left(\psi(b_y+1)-\psi(2)\right)+\frac{\rho_{yx}}{\sqrt{1-\rho_{yx}^2}}\left(\psi(b_x)-\psi(1)\right)\right],$$

$$I_{\mu_y\rho_{yx}} = \frac{-n}{\sigma_y(1-\rho_{yx}^2)}\frac{b_y}{(b_y+2)}\left[\frac{\rho_{yx}^2}{\sqrt{1-\rho_{yx}^2}}\left(\psi(b_y+1)-\psi(2)\right)-\left(\psi(b_x)-\psi(1)\right)\right].$$

4. $$I_{\sigma_y\sigma_y} = \frac{n}{\sigma_y^2}\left\{1+\frac{b_y}{b_y+2}\left[\left(\psi'(b_y+1)+\psi'(2)+\left(\psi(b_y+1)-\psi(2)\right)^2\right)\right.\right.$$

$$+\frac{2\rho_{yx}}{\sqrt{1-\rho_{yx}^2}}\left(\psi(b_x)-\psi(1)\right)\left(\psi(b_y+1)-\psi(2)\right)$$

$$\left.\left.+\frac{\rho_{yx}^2}{1-\rho_{yx}^2}\left(\psi'(b_x)+\psi'(1)+\left(\psi(b_x)-\psi(1)\right)^2\right)\right]\right\},$$

$$I_{\sigma_y\rho_{yx}} = \frac{-n}{\sigma_y(1-\rho_{yx}^2)}\left\{\rho_{yx}-\frac{b_y}{b_y+2}\left[\rho_{yx}\left(\psi'(b_x)+\psi'(1)+\left(\psi(b_x)-\psi(1)\right)^2\right)\right.\right.$$

$$+\frac{1-2\rho_{yx}^2}{\sqrt{1-\rho_{yx}^2}}\left(\psi(b_x)-\psi(1)\right)\left(\psi(b_y+1)-\psi(2)\right)$$

$$-\rho_{yx}\left(\psi'(b_y+1)+\psi'(2)+\left(\psi(b_y+1)-\psi(2)\right)^2\right)\Big]\Big\}.$$

5. $$I_{\rho_{yx}\rho_{yx}}=\frac{n}{(1-\rho_{yx}^2)}\left\{\frac{\rho_{yx}^2}{1-\rho_{yx}^2}+\frac{b_y}{b_y+2}\Big[\psi'(b_x)+\psi'(1)+\left(\psi(b_x)-\psi(1)\right)^2\right.$$

$$-\frac{2\rho_{yx}}{\sqrt{1-\rho_{yx}^2}}\left(\psi(b_x)-\psi(1)\right)\left(\psi(b_y+1)-\psi(2)\right)$$

$$\left.+\frac{\rho_{yx}^2}{1-\rho_{yx}^2}\left(\psi'(b_y+1)+\psi'(2)+\left(\psi(b_y+1)-\psi(2)\right)^2\right)\Big]\right\}.$$

The elements of the information matrix $I_2(\mu_u, \sigma_u, \mu_{v/u}, \sigma_{2.12}, \theta_{vu})$ and the matrix $I_2(\mu_u, \sigma_u, \mu_v, \sigma_v, \rho_{vu})$ are similar to those given above and will not be listed here.

F.2 Student's t

The elements of the information matrix $I_1(\mu_x, \sigma_x, \mu_{y/x}, \sigma_{2.11}, \theta_{yx})$ are as follows (see Tiku et al. (2007)):

1. $$I_{\mu_x\mu_x}=\frac{n}{\sigma_x^2}\left[\frac{r_x+1}{r_x+3}\right],$$

$$I_{\mu_x\sigma_x}=I_{\mu_x\mu_{y/x}}=I_{\mu_x\sigma_{2.11}}=I_{\mu_x\theta_{yx}}=0.$$

2. $$I_{\sigma_x\sigma_x}=\frac{n}{\sigma_x^2}\left[\frac{r_x-3}{r_x+3}\right],$$

$$I_{\sigma_x \mu_{y/x}} = I_{\sigma_{x1}\sigma_{2.11}} = I_{\sigma_x \theta_{yx}} = 0.$$

3. $I_{\mu_{y/x}\mu_{y/x}} = \dfrac{n}{\sigma_{2.11}^2} \dfrac{(r_y+1)}{(r_y+3)}$,

$I_{\mu_{y/x}\sigma_{2.11}} = 0,$

$I_{\mu_{y/x}\theta_{yx}} = \dfrac{n\mu_x}{\sigma_{2.11}^2} \dfrac{(r_y+1)}{(r_y+3)}$.

4. $I_{\sigma_{2.11}\sigma_{2.11}} = \dfrac{n}{\sigma_{2.11}^2}\left[\dfrac{r_y-3}{r_y+3}\right]$, $I_{\sigma_{2.11}\theta_{yx}}$ =0.

5. $I_{\theta_{yx}\theta_{yx}} = \dfrac{n}{\sigma_{2.11}^2}\left\{\dfrac{(r_y+1)}{(r_y+3)} \dfrac{\left[r_x(\sigma_x^2+\mu_x^2)-2\mu_x^2\right]}{(r_x-2)}\right\}$.

The Information Matrix $I_1(\mu_x, \sigma_x, \mu_y, \sigma_y, \rho_{yx})$

1. $I_{\mu_x\mu_x} = \dfrac{n}{\sigma_x^2}\left[\dfrac{r_x+1}{r_x+3} + \dfrac{\rho_{yx}^2}{(1-\rho_{yx}^2)} \dfrac{(r_y+1)}{(r_y+3)}\right]$,

$I_{\mu_x\mu_y} = \dfrac{-n\rho_{yx}}{\sigma_x\sigma_y(1-\rho_{yx}^2)} \dfrac{(r_y+1)}{(r_y+3)}$,

$I_{\mu_x\sigma_x} = I_{\mu_x\sigma_y} = I_{\mu_x\rho_{yx}} = 0.$

2. $I_{\sigma_x\sigma_x} = \dfrac{n}{\sigma_x^2}\left\{1 + \dfrac{r_x-3}{r_x+3} + \dfrac{\rho_{yx}^2}{(1-\rho_{yx}^2)}\left(\dfrac{r_x}{(r_x-2)} \dfrac{(r_y+1)}{(r_y+3)}\right)\right\}$,

$I_{\sigma_x\mu_y} = 0,$

$$I_{\sigma_x\sigma_y} = \frac{-n\rho_{yx}^2}{\sigma_x\sigma_y(1-\rho_{yx}^2)}\left(\frac{r_x}{(r_x-2)}\frac{(r_y+1)}{(r_y+3)}\right),$$

$$I_{\sigma_x\rho_{yx}} = \frac{-n\rho_{yx}}{\sigma_x(1-\rho_{yx}^2)}\left(\frac{r_x}{(r_x-2)}\frac{(r_y+1)}{(r_y+3)}\right).$$

3. $I_{\mu_y\mu_y} = \dfrac{n}{\sigma_y^2(1-\rho_{yx}^2)}\dfrac{(r_y+1)}{(r_y+3)}$, $I_{\mu_y\sigma_y} = I_{\mu_y\rho_{yx}} = 0$

4. $I_{\sigma_y\sigma_y} = \dfrac{n}{\sigma_y^2}\left\{1+\dfrac{r_y-3}{r_y+3}+\dfrac{\rho_{yx}^2}{(1-\rho_{yx}^2)}\left(\dfrac{r_x}{(r_x-2)}\dfrac{(r_y+1)}{(r_y+3)}\right)\right\}$,

$$I_{\sigma_y\rho_{yx}} = \frac{-n\rho_{yx}}{\sigma_y(1-\rho_{yx}^2)}\left\{\frac{r_y-3}{r_y+3}-\frac{r_x}{(r_x-2)}\frac{(r_y+1)}{(r_y+3)}\right\}.$$

5. $I_{\rho_{yx}\rho_{yx}} = \dfrac{n}{(1-\rho_{yx}^2)}\left\{\dfrac{\rho_{yx}^2}{(1-\rho_{yx}^2)}\left(1+\dfrac{r_y-3}{r_y+3}\right)+\dfrac{r_x}{(r_x-2)}\dfrac{(r_y+1)}{(r_y+3)}\right\}$.

The elements of the information matrix $I_2(\mu_u, \sigma_u, \mu_{v/u}, \sigma_{2.12}, \theta_{vu})$ and the matrix $I_2(\mu_u, \sigma_u, \mu_v, \sigma_v, \rho_{vu})$ are similar to those given above and will not be listed here.

APPENDIX G

DATA

G.1 The data for example 5.2.1

Males				Females			
y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4
15	17	24	14	13	14	12	21
17	15	32	26	14	12	14	26
15	14	29	23	12	19	21	21
13	12	10	16	12	13	10	16
20	17	26	28	11	20	16	16
15	21	26	21	12	9	14	18
15	13	26	22	10	13	18	24
13	5	22	22	10	8	13	23
14	7	30	27	12	20	19	23
17	15	30	27	11	10	11	27
17	17	26	20	12	18	25	25
17	20	28	24	14	18	13	26
15	15	29	24	14	10	25	28
18	19	32	28	13	16	8	14
18	18	31	27	14	8	13	25
15	14	26	21	13	16	23	28
18	17	33	26	16	21	26	26
10	14	19	17	14	17	14	14
18	21	30	29	16	16	15	23
18	21	34	26	13	16	23	24
13	17	30	24	2	6	16	21
16	16	16	16	14	16	22	26
11	15	25	23	17	17	22	28
16	13	26	16	16	13	16	14
16	13	23	21	15	14	20	26
18	18	34	24	12	10	12	9
16	15	28	27	14	17	24	23
15	16	29	24	13	15	18	28
18	19	32	23	11	16	18	28
18	16	33	23	7	7	19	18
17	20	21	21	12	15	7	28
19	19	30	28	6	5	6	13

G.2 The children data for example 5.2.2

The variables in the data set are as follows:

x_1 = auditory reception score,

x_2 = visual reception score,

x_3 = visual memory,

x_4 = auditory association,

x_5 = auditory memory,

x_6 = visual association,

x_7 = visual closure,

x_8 = verbal expression,

x_9 = grammatic closure,

x_{10} = manual expression.

Control Group

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
40	32	16	20	38	37	28	43	32	30
35	30	41	44	39	38	32	36	41	27
30	42	48	26	42	34	30	36	43	36
22	27	34	19	40	24	29	14	30	31
21	38	46	28	48	33	28	22	34	26
39	40	47	37	43	40	34	31	49	42
39	39	27	42	36	36	34	47	43	38
22	23	34	16	33	30	33	20	21	30
44	33	43	31	40	40	32	41	40	27
34	34	41	41	51	32	41	46	38	38
30	43	34	46	50	34	32	45	38	33
26	34	32	20	38	20	26	28	28	33
44	42	54	48	54	44	34	52	43	44
36	39	49	24	49	36	35	50	36	36
30	35	32	28	43	32	34	37	39	24
18	25	38	24	32	22	30	8	36	23
27	28	35	25	42	25	36	16	30	24
30	36	42	28	15	26	32	39	28	24
33	16	38	35	51	40	33	40	35	40
26	37	54	32	36	41	38	27	37	32
31	33	33	32	47	37	32	22	36	28
29	31	29	26	38	28	21	27	27	22
34	29	40	26	33	31	30	39	34	26
36	27	34	21	31	27	23	35	36	37
42	40	36	31	41	38	35	41	36	31
32	38	37	42	44	49	32	43	32	40
38	40	40	32	36	41	36	43	41	28

TNT Group									
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
26	38	44	26	37	38	22	28	30	33
31	30	26	23	43	35	28	24	27	33
28	36	36	16	46	28	33	28	28	39
19	43	39	16	46	43	33	28	40	28
31	33	26	29	41	33	23	30	23	24
35	36	43	32	46	22	37	48	34	43
37	35	38	31	44	29	37	41	46	31
41	37	36	29	32	30	32	39	48	27
35	40	34	49	37	34	34	36	34	40
29	19	34	42	32	32	30	40	22	37
18	26	27	26	21	22	26	12	27	26
38	40	27	30	44	40	29	36	42	38
31	23	40	28	37	37	30	45	39	24
26	21	31	29	41	23	24	22	28	26
33	20	44	34	23	37	28	17	39	35
27	36	39	19	28	28	23	38	29	36
22	32	35	26	41	38	25	24	38	26
37	38	36	47	46	38	36	34	32	40
24	40	28	16	36	34	27	24	34	38
11	27	36	19	38	22	23	31	41	35
24	30	30	26	30	36	30	38	35	32
28	14	31	36	48	36	27	38	31	41
17	25	31	35	46	25	30	42	36	27
31	29	34	28	32	31	24	26	29	31
34	38	42	31	43	26	27	27	30	30
41	46	40	39	44	46	36	42	41	40
23	26	37	21	32	34	22	20	31	18

APPENDIX H

THE CASE OF UNEQUAL SAMPLE SIZES

If the sample sizes are different, i.e. $n_1 \neq n_2$, the MML and LS estimators are slightly different. They are given below.

H.1 Generalized Logistic

MML Estimators

1. $\hat{\mu}_x = K_{11} - D_{11}\hat{\sigma}_1$, where

$$K_{11} = \frac{\sum_{i=1}^{n_1} \beta_{1i} x_{(i)}}{m_{11}}, \; D_{11} = \frac{1}{m_{11}} \sum_{i=1}^{n_1} \left(\alpha_{1i} - (b_x + 1)^{-1} \right), \text{ and } m_{11} = \sum_{i=1}^{n_1} \beta_{1i}.$$

2. $\hat{\mu}_u = K_{21} - D_{21}\hat{\sigma}_1$ where,

$$K_{21} = \frac{\sum_{i=1}^{n_2} \gamma_{1i} u_{(i)}}{m_{21}}, \; D_{21} = \frac{1}{m_{21}} \sum_{i=1}^{n_2} \left(\delta_{1i} - (b_u + 1)^{-1} \right), \text{ and } m_{21} = \sum_{i=1}^{n_2} \gamma_{1i}.$$

3. $\hat{\sigma}_1^* = \dfrac{-B_1 + \sqrt{B_1^{\,2} + 4(n_1 + n_2)C_1}}{2(n_1 + n_2)}$, where

$$B_1 = (b_x+1)\sum_{i=1}^{n_1}\left[\left(\alpha_{1i} - (b_x+1)^{-1}\right)\left(x_{(i)} - K_{11}\right) + (b_u+1)\sum_{i=1}^{n_2}\left[\left(\delta_{1i} - (b_u+1)^{-1}\right)\left(u_{(i)} - K_{21}\right)\right]\right]$$

$$C_1 = (b_x+1)\sum_{i=1}^{n_1}\beta_{1i}(x_{(i)} - K_{11})^2 + (b_u+1)\sum_{i=1}^{n_2}\gamma_{1i}(u_{(i)} - K_{21})^2 \ ;$$

adjusting for the bias we get: $\hat{\sigma}_1 = \dfrac{-B_1 + \sqrt{B_1^{\,2} + 4(n_1+n_2)C_1}}{2\sqrt{(n_1+n_2)(n_1+n_2-4)}}$.

4. $\hat{\mu}_{y/x} = \bar{y}_{[.]} - \hat{\theta}\bar{x}_{[.]} - \hat{\sigma}_{2.1}\Delta_1 / m_{12}$, where

$$\Delta_1 = \sum_{i=1}^{n_1}\Delta_{1i}\ ,\quad \Delta_{1i} = (\alpha_{2i} - (b_y+1)^{-1})\ ,$$

$$\bar{y}_{[.]} = \frac{1}{m_{12}}\sum_{i=1}^{n_1}\beta_{2i}y_{[i]}\ ,\ \bar{x}_{[.]} = \frac{1}{m_{12}}\sum_{i=1}^{n_1}\beta_{2i}x_{[i]}\ ;\ m_{12} = \sum_{i=1}^{n_1}\beta_{2i}\ .$$

5. $\hat{\mu}_{v/u} = \bar{v}_{[.]} - \hat{\theta}\bar{u}_{[.]} - \hat{\sigma}_{2.1}\Delta_2 / m_{22}$, where

$$\Delta_2 = \sum_{i=1}^{n_2}\Delta_{2i}\ ,\quad \Delta_{2i} = (\delta_{2i} - (b_v+1)^{-1})\ ,$$

$$\bar{v}_{[.]} = \frac{1}{m_{22}}\sum_{i=1}^{n_2}\gamma_{2i}v_{[i]}\ ,\ \bar{u}_{[.]} = \frac{1}{m_{22}}\sum_{i=1}^{n_2}\gamma_{2i}u_{[i]}\ ;\ m_{22} = \sum_{i=1}^{n_2}\gamma_{2i}\ .$$

6. $\hat{\theta} = K - D\ \hat{\sigma}_{2.1}$, where

$$K = \frac{1}{S}\left[(b_y+1)\sum_{i=1}^{n_1}\beta_{2i}(x_{[i]} - \bar{x}_{[.]})y_{(i)} + (b_v+1)\sum_{i=1}^{n_2}\gamma_{2i}(u_{[i]} - \bar{u}_{[.]})v_{(i)}\right],$$

$$D = \frac{1}{S}\left[(b_y+1)\sum_{i=1}^{n_1}\Delta_{1i}(x_{[i]} - \bar{x}_{[.]}) + (b_v+1)\sum_{i=1}^{n_2}\Delta_{2i}(u_{[i]} - \bar{u}_{[.]})\right],$$

$$S = (b_y + 1)\sum_{i=1}^{n_1} \beta_{2i}(x_{[i]} - \bar{x}_{[.]})^2 + (b_v + 1)\sum_{i=1}^{n_2} \gamma_{2i}(u_{[i]} - \bar{u}_{[.]})^2 .$$

7. $\hat{\sigma}^*_{2.1} = \dfrac{-B + \sqrt{B^2 + 4(n_1 + n_2)C}}{2(n_1 + n_2)}$, where

$$B = (b_y + 1)\sum_{i=1}^{n_1} \Delta_{1i}(y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]}))$$

$$+ (b_v + 1)\sum_{i=1}^{n_2} \Delta_{2i}(v_{[i]} - \bar{v}_{[.]} - K(u_{[i]} - \bar{u}_{[.]})),$$

$$C = (b_y + 1)\sum_{i=1}^{n_1} \beta_{2i}(y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]}))^2$$

$$+ (b_v + 1)\sum_{i=1}^{n_2} \gamma_{2i}(v_{[i]} - \bar{v}_{[.]} - K(u_{[i]} - \bar{u}_{[.]}))^2 .$$

Adjusting for the bias we get: $\hat{\sigma}_{2.1} = \dfrac{-B + \sqrt{B^2 + 4(n_1 + n_2)C}}{2\sqrt{(n_1 + n_2)(n_1 + n_2 - 8)}}$.

8. $\hat{\mu}_y = \bar{y}_{[.]} - \hat{\theta}(\bar{x}_{[.]} - \hat{\mu}_x) - \hat{\sigma}_{2.1}\Delta_1 / m_{12}$.

9. $\hat{\mu}_v = \bar{v}_{[.]} - \hat{\theta}(\bar{u}_{[.]} - \hat{\mu}_u) - \hat{\sigma}_{2.1}\Delta_2 / m_{22}$.

10. $\hat{\sigma}_2 = \sqrt{\hat{\sigma}^2_{2.1} + \hat{\theta}^2\hat{\sigma}^2_1}$.

11. $\hat{\rho} = \hat{\theta}\dfrac{\hat{\sigma}_1}{\hat{\sigma}_2}$.

Least Square Estimators

1. $\tilde{\mu}_x = \bar{x} - (\psi(b_x) - \psi(1))\tilde{\sigma}_1$.

2. $\tilde{\mu}_u = \bar{u} - (\psi(b_u) - \psi(1))\tilde{\sigma}_1$.

3. $\tilde{\mu}_{y/x} = \bar{y} - \tilde{\theta}\bar{x} - (\psi(b_y) - \psi(1))\tilde{\sigma}_{2.1}$.

4. $\tilde{\mu}_{v/u} = \bar{v} - \tilde{\theta}\bar{u} - (\psi(b_v) - \psi(1))\tilde{\sigma}_{2.1}$.

5. $$\tilde{\theta} = \frac{\sum_{i=1}^{n_1}(x_i - \bar{x})y_i + \sum_{i=1}^{n_2}(u_i - \bar{u})v_i}{\sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{i=1}^{n_2}(u_i - \bar{u})^2}.$$

6. $\tilde{\mu}_y = \bar{y} - \tilde{\theta}\tilde{\sigma}_1(\psi(b_x) - \psi(1)) - \tilde{\sigma}_{2.1}(\psi(b_y) - \psi(1))$.

7. $\tilde{\mu}_v = \bar{v} - \tilde{\theta}\tilde{\sigma}_1(\psi(b_u) - \psi(1)) - \tilde{\sigma}_{2.1}(\psi(b_v) - \psi(1))$.

8. $\tilde{\sigma}_1 = \sqrt{\dfrac{(n_1 - 1)s_x^2 + (n_2 - 1)s_u^2}{\left[(n_1 - 1)\left(\psi'(b_x) + \psi'(1)\right) + (n_2 - 1)\left(\psi'(b_u) + \psi'(1)\right)\right]}}$, where

$s_x^2 = \sum_{i=1}^{n_1}(x_i - \bar{x})^2/(n_1 - 1)$ and $s_u^2 = \sum_{i=1}^{n_2}(u_i - \bar{u})^2/(n_2 - 1)$.

9. $\tilde{\sigma}_{2.1} = \sqrt{\frac{\left(\sum_{i=1}^{n_1}\left(y_i - \bar{y} - \tilde{\theta}(x_i - \bar{x})\right)^2 + \sum_{i=1}^{n_2}\left(v_i - \bar{v} - \tilde{\theta}(u_i - \bar{u})\right)^2\right)}{(n_1 - 2)\left(\psi'(b_y) + \psi'(1)\right) + (n_2 - 2)\left(\psi'(b_v) + \psi'(1)\right)}}$.

10. $\tilde{\sigma}_2 = \sqrt{\tilde{\sigma}_{2.1}^2 + \tilde{\theta}^2 \tilde{\sigma}_1^2}$.

11. $\tilde{\rho} = \frac{\tilde{\theta}\tilde{\sigma}_1}{\tilde{\sigma}_2}$.

H.2 Student's t

MML Estimators

1. $\hat{\mu}_x = \frac{\sum_{i=1}^{n_1} \beta_{1i} x_{(i)}}{m_{11}}$,

where $m_{11} = \sum_{i=1}^{n_1} \beta_{1i}$.

2. $\hat{\mu}_u = \frac{\sum_{i=1}^{n_2} \gamma_{1i} u_{(i)}}{m_{21}}$,

where $m_{21} = \sum_{i=1}^{n_2} \gamma_{1i}$.

3. $\hat{\sigma}_1^* = \frac{B_1 + \sqrt{B_1^2 + 4(n_1 + n_2)C_1}}{2(n_1 + n_2)}$, where

$$B_1 = \frac{(r_x+1)}{r_x}\sum_{i=1}^{n_1}\alpha_{1i}x_{(i)} + \frac{(r_u+1)}{r_u}\sum_{i=1}^{n_2}\delta_{1i}u_{(i)} \text{ and}$$

$$C_1 = \frac{(r_x+1)}{r_x}\sum_{i=1}^{n_1}\beta_{1i}(x_{(i)}-\hat{\mu}_x)^2 + \frac{(r_u+1)}{r_u}\sum_{i=1}^{n_2}\gamma_{1i}(u_{(i)}-\hat{\mu}_u)^2 .$$

Adjusting for the bias we get: $\hat{\sigma}_1 = \dfrac{B_1+\sqrt{B_1^{\,2}+4(n_1+n_2)C_1}}{2\sqrt{(n_1+n_2)(n_1+n_2-4)}}$.

4. $\hat{\mu}_{y/x} = \overline{y}_{[.]} - \hat{\theta}\overline{x}_{[.]} + (\hat{\sigma}_{2.1}/m_{12})\sum_{i=1}^{n_1}\alpha_i$, where

$$\overline{y}_{[.]} = \frac{1}{m_{12}}\sum_{i=1}^{n_1}\beta_i y_{[i]} ,\ \overline{x}_{[.]} = \frac{1}{m_{12}}\sum_{i=1}^{n_1}\beta_i x_{[i]}$$

$m_{12} = \sum_{i=1}^{n_1}\beta_i$, and $\beta_i = c_{1[i]}\beta_{2i}$, $\alpha_i = \sqrt{c_{1[i]}}\alpha_{2i}$.

5. $\hat{\mu}_{v/u} = \overline{v}_{[.]} - \hat{\theta}\overline{u}_{[.]} + (\hat{\sigma}_{2.1}/m_{22})\sum_{i=1}^{n_2}\delta_i$, where

$$\overline{v}_{[.]} = \frac{1}{m_{22}}\sum_{i=1}^{n_2}\gamma_i v_{[i]} ,\ \overline{u}_{[.]} = \frac{1}{m_{22}}\sum_{i=1}^{n_2}\gamma_i u_{[i]} ,$$

$m_{22} = \sum_{i=1}^{n_2}\gamma_i$, and $\gamma_i = c_{2[i]}\gamma_{2i}$, $\delta_i = \sqrt{c_{2[i]}}\delta_{2i}$.

Note that under the assumption that $c_1(x) = c_2(u) = 1$, $\sum_{i=1}^{n_1}\alpha_i = \sum_{i=1}^{n_2}\delta_i = 0$.

6. $\hat{\theta} = K + D\ \hat{\sigma}_{2.1}$, where

$$K = \frac{1}{S}\left[\frac{(r_y+1)}{r_y}\sum_{i=1}^{n_1}\beta_i(x_{[i]}-\bar{x}_{[.]})y_{[i]} + \frac{(r_v+1)}{r_v}\sum_{i=1}^{n_2}\gamma_i(u_{[i]}-\bar{u}_{[.]})v_{[i]}\right],$$

$$D = \frac{1}{S}\left[\frac{(r_y+1)}{r_y}\sum_{i=1}^{n_1}\alpha_i(x_{[i]}-\bar{x}_{[.]}) + \frac{(r_v+1)}{r_v}\sum_{i=1}^{n_2}\delta_i(u_{[i]}-\bar{u}_{[.]})\right],$$

$$S = \frac{(r_y+1)}{r_y}\sum_{i=1}^{n_1}\beta_i(x_{[i]}-\bar{x}_{[.]})^2 + \frac{(r_v+1)}{r_v}\sum_{i=1}^{n_2}\gamma_i(u_{[i]}-\bar{u}_{[.]})^2 .$$

7. $\hat{\sigma}^*_{2.1} = \dfrac{B\ +\sqrt{B^2+4(n_1+n_2)C}}{2(n_1+n_2)}$, where

$$B = \frac{(r_y+1)}{r_y}\sum_{i=1}^{n_1}\alpha_i(y_{[i]}-\bar{y}_{[.]}-K(x_{[i]}-\bar{x}_{[.]}))$$

$$+\frac{(r_v+1)}{r_v}\sum_{i=1}^{n_2}\delta_i(v_{[i]}-\bar{v}_{[.]}-K(u_{[i]}-\bar{u}_{[.]})),$$

$$C = \frac{(r_y+1)}{r_y}\sum_{i=1}^{n_1}\beta_i(y_{[i]}-\bar{y}_{[.]}-K(x_{[i]}-\bar{x}_{[.]}))^2$$

$$+\frac{(r_v+1)}{r_v}\sum_{i=1}^{n_2}\gamma_i(v_{[i]}-\bar{v}_{[.]}-K(u_{[i]}-\bar{u}_{[.]}))^2 .$$

Adjusting for the bias we get: $\hat{\sigma}_{2.1} = \dfrac{B\ +\sqrt{B^2+4(n_1+n_2)C}}{2\sqrt{(n_1+n_2)(n_1+n_2-8)}}$.

8. $\hat{\mu}_y = \bar{y}_{[.]} - \dfrac{\hat{\rho}\hat{\sigma}_2}{\hat{\sigma}_1}(\bar{x}_{[.]}-\hat{\mu}_x) = \bar{y}_{[.]} - \hat{\theta}(\bar{x}_{[.]}-\hat{\mu}_x) = \hat{\mu}_{y/x} + \hat{\theta}\ \hat{\mu}_x$.

9. $\hat{\mu}_v = \bar{v}_{[.]} - \dfrac{\hat{\rho}\hat{\sigma}_2}{\hat{\sigma}_1}(\bar{u}_{[.]} - \hat{\mu}_u) = \bar{v}_{[.]} - \hat{\theta}(\bar{u}_{[.]} - \hat{\mu}_u) = \hat{\mu}_{v/u} + \hat{\theta}\ \hat{\mu}_u$.

10. $\hat{\sigma}_2 = \sqrt{\hat{\sigma}_{2.1}^2 + \hat{\theta}^2 \hat{\sigma}_1^2}$.

11. $\hat{\rho} = \hat{\theta}\dfrac{\hat{\sigma}_1}{\hat{\sigma}_2}$.

<u>Least Square Estimators</u>

1. $\tilde{\mu}_x = \bar{x}$.

2. $\tilde{\mu}_u = \bar{u}$.

3. $\tilde{\mu}_{y/x} = \bar{y}_{(.)} - \tilde{\theta}\bar{x}_{(.)}$, where

$$\bar{y}_{(.)} = \frac{\sum_{i=1}^{n_1} c_{1i} y_i}{\sum_{i=1}^{n_1} c_{1i}} \text{ and } \bar{x}_{(.)} = \frac{\sum_{i=1}^{n_1} c_{1i} x_i}{\sum_{i=1}^{n_1} c_{1i}}.$$

4. $\tilde{\mu}_{v/u} = \bar{v}_{(.)} - \tilde{\theta}\bar{u}_{(.)}$, where

$$\bar{v}_{(.)} = \frac{\sum_{i=1}^{n_2} c_{2i} v_i}{\sum_{i=1}^{n_2} c_{2i}} \text{ and } \bar{u}_{(.)} = \frac{\sum_{i=1}^{n_2} c_{2i} u_i}{\sum_{i=1}^{n_2} c_{2i}}.$$

5. $$\tilde{\theta} = \frac{\sum_{i=1}^{n_1} c_{1i}(x_i - \bar{x}_{(.)}) y_i + \sum_{i=1}^{n_2} c_{2i}(u_i - \bar{u}_{(.)}) v_i}{\sum_{i=1}^{n_1} c_{1i}(x_i - \bar{x}_{(.)})^2 + \sum_{i=1}^{n_2} c_{2i}(u_i - \bar{u}_{(.)})^2}.$$

6. $\tilde{\mu}_y = \bar{y}_{(.)} - \frac{\tilde{\rho}\tilde{\sigma}_2}{\tilde{\sigma}_1}(\bar{x}_{(.)} - \tilde{\mu}_x) = \bar{y}_{(.)} - \tilde{\theta}(\bar{x}_{(.)} - \tilde{\mu}_x) = \tilde{\mu}_{y/x} + \tilde{\theta}\bar{x}$.

7. $\tilde{\mu}_v = \bar{v}_{(.)} - \frac{\tilde{\rho}\tilde{\sigma}_2}{\tilde{\sigma}_1}(\bar{u}_{(.)} - \tilde{\mu}_u) = \bar{v}_{(.)} - \tilde{\theta}(\bar{u}_{(.)} - \tilde{\mu}_u) = \tilde{\mu}_{v/u} + \tilde{\theta}\bar{u}$.

8. $\tilde{\sigma}_1 = \sqrt{\frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_u^2}{\left((n_1 - 1)\frac{r_x}{r_x - 2} + (n_2 - 1)\frac{r_u}{r_u - 2}\right)}}$, where

$s_x^2 = \sum_{i=1}^{n_1}(x_i - \bar{x})^2 / (n_1 - 1)$, and

$s_u^2 = \sum_{i=1}^{n_2}(u_i - \bar{u})^2 / (n_2 - 1)$.

9. $$\tilde{\sigma}_{2.1} = \sqrt{\frac{\left[\sum_{i=1}^{n_1} c_{1i}\left(y_i - \bar{y}_{(.)} - \tilde{\theta}(x_i - \bar{x}_{(.)})\right)^2 + \sum_{i=1}^{n_2} c_{2i}\left(v_i - \bar{v}_{(.)} - \tilde{\theta}(u_i - \bar{u}_{(.)})\right)^2\right]}{\left[(n_1 - 2)\frac{r_y}{r_y - 2} + (n_2 - 2)\frac{r_v}{r_v - 2}\right]}}.$$

10. $\tilde{\sigma}_2 = \sqrt{\tilde{\sigma}_{2.1}^2 + \tilde{\theta}^2 \tilde{\sigma}_1^2}$.

11. $\tilde{\rho} = \dfrac{\tilde{\theta}\tilde{\sigma}_1}{\tilde{\sigma}_2}$.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Qumsiyeh, Sahar
Nationality: Palestinian
Date and Place of Birth: 29 April 1971, Bethlehem
Marital Status: Single
Phone: +90 538 3176505
email: saharq@yahoo.com

EDUCATION

Degree	**Institution**	**Year of Graduation**
MS	Brigham Young University, USA	1997
BS	Bethlehem University, Palestine	1993
High School	Talitha Kumi, Palestine	1987

WORK EXPERIENCE

Year	**Place**	**Enrollment**
2003-2004	AAUJ, Palestine	Statistics Instructor
2002-2003	BU, Palestine	Mathematics Instructor
2001-2002	BYU, USA	Teaching Assistant
1998-1999	BU, Palestine	Mathematics Instructor
1996-1998	Palestine Polytechnic Inst.	Mathematics Instructor

FOREIGN LANGUAGES

Fluent English, Arabic, and Turkish

PUBLICATIONS

"Partitioning the Degrees of Freedom for Age-Period-Cohort Model Analysis," MS Thesis, Brigham Young University, 1997

HOBBIES

Reading, Computer Programming, Swimming.